MAXIMIZING INTELLIGIBILITY OF ENGLISH LANGUAGE TEXT SETTING

and

HEAVEN'S GATE

by

DAVID WOLFSON

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Music

Written under the direction of

Nancy Rao

And approved by

____

____

____

____

New Brunswick, New Jersey

January, 2018

ABSTRACT OF THE DISSERTATION

Maximizing Intelligibility Of English Language Text Setting

by DAVID WOLFSON

Dissertation Director:

Nancy Rao

Is there anything a composer can do to maximize the chances that a setting of English text for the classically trained voice can be understood by audiences? Intelligibility can be a problem for composers of English-language opera and art songs, despite the best efforts of their singers. The existing literature from composers and composition teachers on the subject of maximizing intelligibility is negligible. I draw on recent research by Lauren Collister and David Huron, as well as Nicole Scotto di Carlo (among others), who are beginning to shed some light on the ways in which song is heard differently than speech, on what aspects of classical singing can obscure listener comprehension, and on some specific ways in which different aspects of text setting can affect the text's intelligibility, and take an interdisciplinary approach to extending their work, introducing relevant ideas and results from the fields of phonetics and psycholinguistics to interrogate compositional practices.

I suggest that these and other investigations point toward the possibility of a more comprehensive analytical toolkit for composers, singers, analysts, and others interested in the intelligibility of classically trained singers.

ACKNOWLEDGEMENTS

I would like to thank:

•the members of my dissertation committee: Robert Aldridge, Steven Kemper, Steven Rosenhaus, and especially my advisor, Nancy Rao;

•all those who generously took the time to make suggestions or explain the intricacies of disciplines I knew nothing about: Lauren Ackerman, John T. Hale, Erich Heins, Joanne Sydney Lessner, Brian Roark, Martha Sullivan, Lisabeth Miller Kettledon, Lauri Aguirre, Betsy Podsiadlo, Giulia Utz, and especially Experience Bryon;

•the cast, staff, musicians and crew of Hartford Opera Theatre's production of *Heaven's Gate*;

•Charles Fussell, Robert Aldridge, and Experience Bryon for their suggestions and insights in the writing and composition of *Heaven's Gate*;

•and most especially Lynn Wichern, without whom none of this would have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF MUSICAL EXAMPLES

Introduction

In their 2015 paper "Catching the Lyrics: Intelligibility in Twelve Song Genres," music theorists Nathaniel Condit-Schultz and David Huron compare the intelligibility of sung English words across a spectrum of different musical categories. Their results show that the intelligibility of lyrics (text), or the lack thereof, is a problematic aspect of most types of vocal music; the average percentage of lyrics understood by their testers across all genres was 72%—meaning that more than a quarter of all those words were unintelligible. The percentage for classical music, however, was much lower—the lowest of all the genres surveyed—at 46%. More than half of the English words sung by classically trained singers were unintelligible to their English-speaking audiences.[1] The use of supertitles for English-language operas in English-speaking countries attests to the widespread awareness of this issue, as does the common practice of printing the texts even of English-language art songs in recital programs.[2]

While not every composer has intelligibility of text as a desideratum—and even those who do may not consider it a universal or primary goal—there are and have been numerous complaints from critics and audiences about the intelligibility of classically sung English in contexts where it seems likely that audience comprehension of the words contributes heavily to the artistic goals of the piece. Here is a sampling.

From an opera review in 1909:

---

[1] Nathaniel Condit-Schultz and David Huron, "Catching the Lyrics: Intelligibility in Twelve Song Genres," Music Perception: An Interdisciplinary Journal 32 (2015): 476.

[2] While many of these issues may also afflict text sung in other languages as well (particularly for those reasons discussed in Chapters 2, 3, and 5 of this paper), I have chosen to discuss the setting of English language text exclusively, as English is my native language. Generalization of this approach to include other languages is certainly possible.

> It was somewhat discouraging to the advocates of opera in English that these English-speaking singers had so little success in making their words understood. Mr. Whitehill achieved the best results in this way; there were long passages delivered by the other chief singers in which little or nothing of the text was intelligible."[3]

From an opera broadcast review in 2013:

> This brings me to my latest review of Gioachino Rossini's *The Barber of Seville*. An English-language adaptation of the Pesaro-born composer's delightful nineteenth-century masterwork, it was broadcast on December 22, in director Bartlett Sher's crowd-pleasing production. The main problem here wasn't so much the non-colloquial Italian as it was unintelligible English.[4]

A 2001 letter to the editor of The Daily Telegraph in London:

> The Daily Telegraph received a letter on 24 March asking whether anybody but that paper's opera critic Rupert Christiansen has ever understood more than five consecutive words sung by an operatic soprano in English.[5]

As the plaintive writer of that letter suggests, far from being seen as restricted to certain unfortunate performances, the phenomenon is considered widespread. The supertitles or printed texts offered as solutions to the problem are at best stopgaps. As one critic put it:

> In opera, understanding words is always a problem with an unfamiliar work. The supertitles above the stage are difficult to read and not much help. (Besides, though reading is a wonderful thing, no one comes to the theater to do it.)[6]

Similar complaints have been registered about art song:

---

[3] "At the Opera," review of *The Pipe of Desire*, *The New Music Review and Church Music Review* 97 (1909)
[4] Josmar Lopes, "Opera Review: 'The Barber of Seville' in English – Shave and a Haircut, Two Bits," Reviews by Josmar Lopes, https://josmarlopes.wordpress.com/2013/01/02/opera-review-the-barber-of-seville-in-english-shave-and-a-haircut-two-bits/ (accessed 2/29/2016)
[5] "English opera sparks debate," BBC News, http://news.bbc.co.uk/2/hi/entertainment/1245186.stm (accessed 2/29/2016)
[6] Marion Hunter, "'Wreckers,' rare opera penned by woman, opens at Bard," columbiapaper.com, https://www.columbiapaper.com/2015/07/online-only-theater-review-wreckers-rare-opera-penned-by-woman-opens-at-bard/ (accessed 2/29/2016)

> Intelligibility of text is an important component of art song…The most recent (traditional) art song recital I attended was comprised wholly of songs in English, yet the majority of the audience buried their faces in the program as soon as the concert began; it seemed like a reflex response.[7]

Particularly in the case of opera, though, the issue is major. New Yorker contributor James B. Stewart included unintelligibility of English-language operas among the "huge challenges" facing opera in a 2015 article about the Metropolitan Opera.[8] And Philadelphia writer Tom Purdom bemoaned the lost opportunity for connecting with new audiences in a 2013 production of *Owen Wingrave:*

> Unfortunately, the production made no attempt to accommodate all the people who thought they were going to hear a musical dialogue on one of the greatest conundrums facing our species. The drama on the stage was as unintelligible as a Sanskrit text.[9]

Given the continually precarious state of opera in the current climate, an increase in intelligibility could be a very real contribution to the continued viability of the art form.

Aims of this paper

What can the composer who wants to create an intelligible setting of English text do to maximize the chances that it will be understood? This paper offers a new approach to that question.

---

[7] Corey Dargel, "More Song, Less Art(ifice): The New Breed of Art Song," newmusicbox.org, http://www.newmusicbox.org/articles/More-Song-Less-Artifice-The-New-Breed-of-Art-Song/ (accessed 2/29/2016)
[8] James B. Stewart, "A Fight at the Opera," newyorker.com, http://www.newyorker.com/magazine/2015/03/23/a-fight-at-the-opera (accessed 2/29/2016)
[9] Tom Purdom, "Classical promoter, cure thyself: A cautionary tale," broadstreetreview.com, http://www.broadstreetreview.com/music/music_marketing_missed_opportunities (accessed 2/29/2016)

In Section One (Chapters 1 and 2) I examine the resources currently available to the composer or analyst interested in this subject. Chapter 1 reviews the sparse existing literature on intelligibility of musical setting for the classical voice, and Chapter 2 reviews both the differences from spoken language that make classically sung English so hard to understand and the spate of recent empirical research investigating the questions of how exactly these differences manifest themselves.

In Section Two (Chapters 3, 4 and 5) I outline an interdisciplinary approach in which inquiries and results from disciplines not usually considered by composers or music theorists can contribute to our understanding of what makes a set text more or less intelligible. Chapter 3 explores questions from selected areas of cognitive science to look at the ways in which the functioning of human memory and processing affect the intelligibility of sung text; in certain ways, it seems that understanding language and attending to music are mutually exclusive tasks. Chapter 4 applies phonetics to examine the ways in which phonemes contribute to the difficulty of processing of language at the level of the word; specifically, it explores the issue of how we distinguish among ambiguous interpretations of heard phonemes, and what happens when we are unable to. Chapter 5 incorporates findings from psycholinguistics and computational linguistics to examine the ways in which words contribute to the difficulty of processing of language at the level of the sentence. Models of sentence comprehension indicate that syntactically and lexically unexpected words are going to be more difficult to understand in spoken or read contexts, let alone when sung.

In Section Three (Chapters 6 through 13), applying all of these disciplines' approaches to intelligibility, I construct a methodology for analysis of the likely intelligibility of a sung text, and apply it. Chapter 6 combines the results outlined in the previous four chapters into an analytical method, noting that it could also be used prescriptively by composers wishing their settings to be maximally intelligible. In Chapters 7 through 13 I employ this methodology to case studies of selections from art song, opera, operetta, and that portion of the musical theatre literature written to be performed by opera singers, affording a look at how composers of English-language art song and music theatre (to use the umbrella term) have grappled with the issues of intelligibility, and how their work compares with the prescriptions of the method outlined in Chapter 6.

It is my hope that this work will supply some practical guidance to those composers, librettists and translators of music for the classically trained voice for whom the intelligibility of the sung text is important—not to mention some relief to those singers who are asked to communicate inherently unintelligible text settings. While there are certainly other values and aspects to vocal music, any of which may in a given situation win out over text comprehension in a composer's intent, the wish to understand the words they are hearing is widespread among audiences, and the future health of English-language opera and art song can only be helped by an increase in their intelligibility.

Chapter 1: The State of the Art

Is there any guidance available for the composer who wishes to set text for the classical voice as intelligibly as possible? Is there any source to help determine whether a given text, a given phrase or word, will be difficult to understand or easy? If that were to be determined, is there any advice available about how to treat such words or phrases musically to maximize their intelligibility? The answer to all of these questions is: not much. This chapter, in which I critically review the existing literature devoted specifically to these questions, will be a short one.

Virgil Thomson

Virgil Thomson's 1989 collection of musings and essays *Music With Words: A Composer's View* covers territory ranging from the durational extensibility of vowels to the nature of opera. His stated aim is "to establish a method of operation for vocal writing in English or in American."[10] The importance to this enterprise of intelligibility turns up on the first page: "…it is pleasanter to understand what music is saying than merely to enjoy the sound that it makes."[11]

Thomson deals with the setting of text in the early chapters of the book. His advice consists of the following:

> As a reminder that the art of putting English to music is largely a matter of not disturbing the fixed elements, let me repeat that the attributes of speech-sound are: stress, or accentuation, which in English is invariable; cadence, which is extremely variable—but only within the limits of the third attribute, quantity, since certain sounds are considered extensible and others not. Accents in English cannot be

---

[10] Virgil Thomson, Music With Words: A Composer's View (New Haven: Yale University Press, 1989): 3.
[11] Ibid: 1.

> changed without changing the meaning. Cadence can be widely varied to illustrate meanings or to intensify them, but only where the phonemes, or units of speech-sound, are in themselves extensible.[12]

(He defines cadence, in this sense, as "variations in pitch, the up-and-downness of a phrase or word-grouping"[13]. The more usual term is prosody.)

He goes on to note that words have natural lengths determined not only by the lengths of their vowels, but of their consonants, such that *pope* is a shorter word than *home,* and that one should "allow a bit of time for the long ones and…not expect the short ones to be held."[14] While he mentions and categorizes unstressed vowels ("mutes"), rhoticized vowels ("*r* vowels"), and diphthongs, and their relative lengths in spoken English, he has no further advice about how to set them.

Thomson makes a more subtle contribution by introducing the concept of *word-groups*: words that form a unit of pronunciation, and cannot be separated without disrupting their meaning. This insight is based on the fact that, in spoken language, separations do not occur at the boundaries between words; in general, we run words together into streams of sound. He advocates analysis of a text to be set into word groups (and, implicitly, the setting of the text so that word-groups are not disrupted).[15]

His advice pertaining to pitch is to make a list of significant words in the text, and set them in the vocal line in the upper half of the middle register.[16] He also has a bit of guidance about the contribution of instrumental accompaniment to vocal intelligibility:

---

[12] Ibid: 9.
[13] Ibid: 4.
[14] Ibid: 10.
[15] Ibid: 16-17.
[16] Ibid: 28.

rhythmic independence of the accompaniment may interfere with the projection of the words unless the vocal line is simple.[17]

As welcome as these suggestions are, they are short on specifics and leave large areas of inquiry untouched. Particularly as to the question of knowing whether a given text is suitable for setting, he shrugs: "Instinct and experience must be the guide."[18] I argue in later chapters that there are ways to put the process of text selection on a less instinctual basis.

Paul Barker

Composer and musical theatre performance pedagogue Barker's 2004 book *Composing for Voice* aims to be an encyclopedic resource on the subject. On the matter of intelligibility, however, it falls far short of that. His Chapter 3, "Voice/Text/Music," introduces the issues surrounding intelligibility lucidly enough, nodding to the type of audience complaint mentioned in the Introduction to this paper and acknowledging that the composer may be as much at fault as the singer.[19] Like Thomson, he notes the constraints for the composer of the language's innate features:

> If linguistic comprehension is an objective… then there are a series of consequences affecting all rhythms and pitches, derived from a study of how the words might be made intelligible in a musical context, with an appropriate voice.[20]

However, if he has made this sort of study, he does not offer its results in this book. The counsel on offer here is even less specific than in the Thomson:

---

[17] Ibid: 39.
[18] Ibid: 74.
[19] Paul Barker, *Composing for Voice*, (New York: Routledge, 2004): 45.
[20] Ibid.

> If a fluid line with intelligible text is an objective, composers should develop sensitivity to the most appropriate syllables for the pitches within a phrase, particularly at extremes of range. Appropriateness may be a combination of many factors, including emotive words, syllabic stress, and the category of vowel sound.[21]

How a composer might develop that sensitivity is not addressed. There are no examples given that directly illustrate these factors.

While the rest of the chapter offers a great deal of descriptive vocabulary—there are sections on vowels and consonants, with additional breakouts on diphthongs and sibilants—the subject of intelligibility does not come up again.

John Halle

Composer and music theorist Halle, in an online posting of a book in progress called *Text, Tune and Metrical Form,* goes into a great deal of detail on one aspect of the issue. He takes as his starting point the idea of matching rhythmic with metric stresses, the procedure implied by Thomson's dictum that stresses in English are invariable and should not be "disturbed." He goes on to create a formalism for how this should be done, drawing on a previous paper (with Fred Lardil) called "A Generative Textsetting Model," itself an outgrowth of Lerdahl and Ray Jackendoff's foundational *A Generative Theory of Tonal Music.* Most interesting is the notion that different linguistic groupings—clitic group, phonological phrase, intonational phrase, and utterance—should align with musical groupings for a good, well-formed text setting. (The parallel with Thomson's word-groups is obvious.) Note that his criterion is "acceptability," however—the sense

---

[21] Ibid: 56.

that it feels well-formed and normative—rather than intelligibility when sung.[22] Halle deals almost exclusively with rhythm: "the intuitions which govern English text setting appear to be largely rhythmic and do not implicate pitch."[23] He mentions in passing that musical groupings (of the type that should be aligned with linguistic groupings) are usually separated by long notes, but can also be marked by changes in dynamic, register or some other quality, but does not go on to explore the compositional ramifications of this idea.

While there are other sources available to the composer of music for the classically trained voice (a web search will turn up tutorial papers and how-to videos), the advice proffered in these echoes that of Thomson: align metric stresses with musical stresses; don't expect words set to notes above the staff to be intelligible.

<u>The view from musical theatre</u>

What about the world of musical theatre, where the words have far more status than they sometimes seem to in opera? The preferred vocal technique in contemporary musical theatre has fewer obstacles to comprehension than does that in opera (not to mention the assistance of amplification and sound mixing); but the stakes for comprehension are much higher, given the genre's historical expectations of narrative through song. Cohen and Rosenhaus' *Writing Musical Theater*, a recent addition to the small pool of books by composers on the subject, offers these tidbits for musical theatre songwriters: "eschew such 'poetic' practices as inverting word order;" avoid archaic

---

[22] John Halle, *Text, Tune and Metrical Form,* http://johnhalle.com/musical.writing.technical/technical.htm (accessed 2/18/16).
[23] Ibid: 64.

language; match linguistic stresses to musical stresses; avoid combinations of words that can be heard in more than one way (phrasal homophones), such as "a day of festival/a day, a festival;" avoid combinations of sounds that are inherently difficult to sing, such as densely packed consonants or frequent sibilants; and "Vowel sounds in certain ranges can also be a problem."[24]

This, of course, is advice to the lyricist. For the composer, we have: avoid setting closed vowels on high notes; and, if the lyricist can't avoid the pitfalls mentioned above, it's up to the composer to ameliorate the situation (e.g., separate identical adjacent consonants with a rest).[25]

Here, in the admonition to avoid phrasal homophones, is the first acknowledgement that certain combinations of words may simply be harder to understand when sung than others. I will be examining this idea further in Section Two, which deals with approaches to intelligibility from other disciplines.

With the exception of Halle's unfinished opus, all of the above advice could be considered conventional wisdom, part of the body of knowledge a composer will absorb from lessons and from listening to and working with singers over the course of a career. It is valuable enough information. But this body of knowledge is beginning to be supplemented, and detail added, by musicological experiment and research undertaken only in the last fifteen years. Chapter Two will put that literature into context with the well-understood differences from spoken language that make classically sung English so difficult to comprehend.

---

[24] Allen Cohen and Steven L. Rosenhaus, *Writing Musical Theater,* (New York: Palgrave MacMillan, 2006): 113-115.
[25] Ibid: 146.

Chapter 2: Why Can't We Understand Sung English?

For a 2008 paper, music theorists Lauren Collister and David Huron asked groups of volunteers to listen to recordings of the same phrases spoken and sung and transcribe what they heard. The experiment was designed to eliminate contextual effects, isolating the recorded words outside of any syntactic context. The reduction in intelligibility of sung English was thus quantified for the first time: sung words were roughly one-seventh as intelligible as the spoken ones. [26] But why are classically sung words so hard to understand?

Music psychologists Philip Fine and Jane Ginsborg have divided the factors into four groups: performer-related, listener-related, environment-related and words/music related.[27] This thesis is concerned with the latter category, the contribution of the words and music themselves to the intelligibility (or lack thereof) of a given piece of classical vocal music. So, much as a physicist assumes a perfect sphere or a frictionless plane (in order to simplify and isolate the phenomena being studied), we will assume for the remainder of this paper that we are discussing the case of an attentive, alert and competent listener, in an acoustically friendly hall, listening to a singer with excellent diction.

---

[26] Lauren B. Collister and David Huron, "Comparison of Word Intelligibility in Spoken and Sung Phrases," *Empirical Musicology Review* 3 (2008): 109. This should be taken as an approximate figure; the authors note that the lack of linguistic context would make the words less intelligible overall (a topic I will be investigating in Chapter 4). I note that, based on the examples in the paper, they did not make a point of aligning linguistic with metric stresses in the sung portion of their experimental material, which would also potentially decrease intelligibility.

[27] Philip A. Fine and Jane Ginsborg, "Making myself understood: perceived factors affecting the intelligibility of sung text," *Frontiers in Psychology* 5 (2014): 1.

Excellent diction, however, does not in itself make a singer's words intelligible. There are significant differences between spoken and sung language that interfere, no matter how good the singer's articulation. This chapter will review the well-understood factors that make sung phonemes more difficult to recognize than their spoken counterparts (as well as other differences between speech and song), along with the results of Collister and Huron and their colleagues' investigations into musical and textual elements that combine to affect the intelligibility of sung English.

Vowels

Vowels sung above a certain pitch tend to lose their distinctive identities, making intelligibility a particular issue for sopranos. To understand why this is so, we need to introduce the acoustic concept of *formants.* Formants are resonant frequency bands that come from the shape of the vocal tract (being reinforcements of harmonics of the fundamental pitch).[28] Vowels are identified by the first two formants, while the formants above contribute to the sound of the individual voices. Formants and pitch are independent; a singer can change the pitch and keep the vowel constant, and vice versa. However, when the fundamental pitch is higher than the normal lowest formant of a vowel, the fundamental itself is substituted for that lowest formant, changing the perceived vowel.[29] The first formants of English vowels range from about 250 to 750 Hz (roughly B3 to F#5), all within the normal range of the female voice (and many male ones). Thus,

---

[28] This type of formant is not to be confused with the "singer's formant," about which more shortly.
[29] This effect is acoustic, and may differ from perceptions of the effects of pitch on anatomy (and vice versa) in many vocal pedagogies.

> "when the fundamental has a frequency appreciably higher than the first formant frequency of a given vowel, this fundamental tends to impose the perception of one of the vowels that would have a first formant at that frequency. An extreme example is singing vowels on G792—a fundamental frequency nearly equivalent to the normal first formant frequency of [ɑ]. The result, here, is very poor intelligibility for all vowels except [ɑ], with the other vowels being largely interpreted as [ɑ] under the conditions of our listening tests."30

The effect is not limited to high notes for sopranos. Female singers increase F_1 (the first formant) as they increase F_0 (the fundamental) for vowels whose first formant is lower in pitch than the fundamental. Thus, we would expect that "as the fundamental rises, "i" will begin to sound like "I" and then "Ɛ;" "I" will begin to sound like "Ɛ" and then "æ;" and "Ɛ" will begin to sound like "æ." This indeed happens."[31] Since the characteristic formants of the vowels mentioned here range from 240 to about 600 Hz, this effect is in play in most of the range of most classically trained female singers, and it seems likely that it would be relevant to at least some male voices as well. /i/ begins to be interpreted as /I/ at 587 Hz (D5) and as /Ɛ/ at 740 Hz (F#5); /I/ begins to sound like /Ɛ/ at D5 as well; /Ɛ/ begins to be interpreted as /æ/ as low as 370 Hz (F#4).[32]

---

[30] John Howie and Pierre Delattre, "An Experimental Study of the effect Pitch on the Intelligibility of Vowels," *NATS Bulletin* 18 (1962): 6-9. There are, of course, exceptions. In an experiment by William Triplett, a soprano sang a series of vowels on a high C. They all sounded like [ɑ], as expected, expected for the vowel [i]. Upon examination of the spectrogram for that vowel, it turned out that she had (unconsciously) emphasized the 3rd harmonic at the beginning of the sound, and for the listeners this perception had carried through the duration of the note. (Triplett 1967: 6, 8, 50). Strategies of this sort may be used by some singers in some situations; another was outlined by Lloyd A. Smith and Brian L. Scott, who asked a soprano to sing high notes with a raised larynx, and discovered that this made her vowels more distinguishable, although they noted drily that the resulting sound "was certainly not one to be recommended for singers to produce." (Smith and Scott 1980: 1795)

[31] Martha S. Benolken and Charles E. Swanson, "The effect of pitch-related changes on the perception of sung vowels," *Journal of the Acoustical Society of America* 87 (1990): 1782. Benolken and Swanson also expanded on Howie and Delalltre's findings: although /u/, /o/ or /Ɛ/ are almost universally interpreted as /ɑ/ at high pitches, /i/ and /I/ were interpreted as most of the vowel choices they gave their listeners. (1784)

[32] Ibid: 1783

All of the investigations cited above were into individual, pure vowel sounds. Vowels do not necessarily come in splendid isolation, however. In Collister and Huron's experiments, diphthongs were more easily identified than monophthongs (75% vs. 69%),[33] possibly because the sound of the transitions between the constituent vowels helped identify them (see the section on transitions below).

The Singer's Formant

The "singer's formant," previously mentioned, is a constant resonant frequency, independent of both pitch and vowel, peculiar to classically trained voices, and more pronounced in men than in women (researchers seem to disagree on whether sopranos have singer's formants); it is what enables these singers to project over an orchestra. However, its salubrious effect on audibility is balanced by a deleterious one on intelligibility: the singer's formant masks *formant transitions*, making consonants harder to distinguish.[34]

Transitions

A *formant transition* is the change in a vowel's formant as it goes to or from a consonant before or after it. Why should formant transitions matter to intelligibility? As voice scientist Ingo Titze puts it, "much of the verbal message is transmitted in transitions from one articulatory state to another… Articulators in continual motion, with precision timing amongst each other, provide a framework of acoustic cues that are of

---

[33] Collister and Huron, 8-9.
[34] Nicole Scotto Di Carlo, "Effect of multifactorial constraints on intelligibility of opera singing," *Journal of Singing* 63 (2007): 446.

themselves dynamic."[35] In other words, vowels and consonants make each other intelligible by creating each other's context.

In addition to the "singer's formant" masking effect described by Nicole Scotto DiCaprio, the formant transitions of classically trained singers may be quicker. Voice scientists Johann Sundberg and Camilla Romedahl, in a head-to-head comparison of intelligibility between musical theatre and opera singers masked by noise, discovered that the opera singers' "formant transitions in consonants were much quicker than those of the MT [musical theatre] singers, and this may have reduced text intelligibility."[36]

Consonants

Consonants are also affected by classically trained vocal technique. Besides the fact that the requirements of clear diction and the legato line are essentially incompatible, leading singers to underarticulate,[37] and the fact mentioned above that transitions from vowel to consonant or vice versa seem to happen faster in singing, certain consonants suffer just by being part of a sung line.

> As the intensity and fundamental frequency of a sung vowel increase, the intensity of the adjacent consonant decreases… This leads to confusions in the identification of those consonant categories for which noise intensity and frequency are the main acoustic cues.[38]

These consonant categories are the unvoiced fricatives and (to a lesser extent) the unvoiced plosives, a not inconsiderable group. However, all consonants decline in

---

[35] Ingo Titze, "Why is the Verbal Message Less Intelligible in Singing than in Speech?" *NATS Bulletin* 38 (1982): 37.
[36] Johann Sundberg and Camilla Romedahl, "Text Intelligibility and the Singer's Formant—A Relationship?" *Journal of Voice* 23 (2009): 544.
[37] Di Carlo: 563.
[38] Ibid: 449.

intelligibility as the vocal line climbs in register (although some, like the nasal consonants /ɲ/, /m/, /n/, and the liquid consonant /l/, are more resistant to the effect than others).[39]

Collister and Huron kept a tally of the phonemes that were most vulnerable to mishearing. Liquids (/l/, /ɹ/)had the fewest errors. Nasals (/m/, /n/, /ŋ/) and stops had a greater number. Voiced stops (/b/, /d/, /g/) had a greater number of errors than voiceless. There were also consonant insertions, mostly at the beginnings and ends of words, and deletions, particularly of liquids and nasals, when they seemed to merge with the neighboring vowel.[40]

Vibrato

Vibrato, a characteristic feature of classically trained singing, also masks the formant frequencies and the direction of the formant transitions. This makes both vowels and consonants more difficult to identify.[41] The pitch variation in vibrato does, in the course of its cycle, reinforce formant resonances, but according to acoustic researcher Johan Sundburg this effect does not outweigh the other.[42]

In a later study (in which they were joined by Randolph Johnson), Collister and Huron reported no difference in intelligibility between the two operatic voices and the four musical theatre voices that were used in their trials.[43] (They had hypothesized that,

---

[39] Ibid: 450.
[40] Collister and Huron: 116-117.
[41] DiCarlo: 563.
[42] Johan Sundberg, "Acoustic and Psychoacoustic effects of vocal vibrato," *STL-QPSR* 35 (1994): 62.
[43] Randolph Johnson, David Huron & Lauren Collister, "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," *Empirical Musicology Review* 9 (2014): 15.

because "in musical theatre…greater priority is given to conveying the words to the audience," the musical theatre singers would be more intelligible.)[44] However, while both of the descriptors in question can cover a vast array of training styles and techniques, they described the difference only as one of the amount of vibrato used. Furthermore, I suspect that the extremely artificial circumstances of recording the stimuli for their experiments is sufficiently different from the actual experience of performance for both groups that this result is not particularly meaningful.

Thus far we've discussed the sounds of the phonemes themselves, and how they and the transitions between them might differ, in the course of a sung line, from the same text spoken, effects that certainly contribute to the difference in intelligibility between spoken and sung language. But there are other differences between language spoken and language sung.

Durational cues

The relative lengths of syllables and consonants are, in normal speech, cues that help us recognize the phonemes that make them up. Vowels are longest in open syllables, shorter in syllables closed by a voiced consonant, shortest in syllables closed by a voiceless consonant. Vowels are longer in stressed syllables. Vowels get shorter as syllables are added to a word.[45] [46] Needless to say, these cues are missing in most sung speech, where the rhythms of the music replace the rhythms the same words would have

---

[44] Ibid: 5.
[45] William F. Katz, *Phonetics for Dummies,* Hoboken: John Wiley & Sons, 2013: 121.
[46] This is the background to Virgil Thomson's observation that words have inherent lengths, discussed in Chapter 1.

when spoken (certain types of recitative excepted). Moreover, consonants are generally slightly shorter when sung than when spoken; voiced consonants, normally quicker than their unvoiced counterparts, are even more brief as part of a sung line.[47]

The effect of the slow rates of delivery of some vocal lines on intelligibility is an unexplored question.[48] But Collister and Huron found that melismatic melodies, in which the rate of delivery is presumably slower than that of syllabic settings, gave twice the rate of error of all their other trials (their melismas ranged from five notes to twenty-two).[49]

In their later study with Johnson, they confirmed that words in syllabic settings are more easily identified than words in melismatic settings. They also found that bisyllabic words were more easily identified than monosyllabic words in melismatic settings, but were not certain whether that was because, in the design of their experiment, the length of melisma per syllable was shorter.[50] (As we shall see in Chapter Five, there is a reason why bisyllabic words are easier to identify in general than monosyllabic words.)

Phonetics rules

There are several rules of English language phonetics—processes of alteration of phonemes in certain contexts, naturally produced and understood by native speakers—that are subject to change when English is sung. In spoken English, nasal sounds and liquids become syllabic at the ends of words. Vowels at the beginnings of words or phrases are preceded by a glottal stop. Approximants become partially devoiced after

---

[47] Di Carlo: 559-560.
[48] Condit-Shultz and Huron (2015) note that an obvious next question is how the rate of singing (in syllables per second) affects intelligibility.
[49] Collister and Huron: 119.
[50] Johnson, Huron & Collister: 10-11.

aspirated stops.[51] Voiceless stops at the beginnings of syllables are associated with longer voice onset time than voiced stops.[52] All of these are either difficult to do when singing, or would usually be jettisoned in the pursuit of the legato line.

Prosody

The normal pitch content of spoken English—its "linguistic prosody"—is replaced by the pitch content of the melodic setting when sung. This spoken pitch content is not just a conveyor of subtext; it also conveys grammatical information on the level of the sentence (as in the intonational difference between questions and statements).[53]

Confirming the insights of Thomson, Barker and others, Johnson, Huron and Collister found that words set with the metric stress matching the syllable stresses were more easily identified than with those stresses mismatched. Note that they used only bisyllabic words in this experiment.[54]

The subtler questions of how (if at all) melodic contour affects intelligibility, and how other forms of musical stress besides metric interact with other levels of linguistic stress besides word stress, are as yet unexplored.

Familiarity

Some words are simply more intelligible than others when sung. Johnson, Huron and Collister found that familiar words were more intelligible (more easily identified)

---

[51] Katz: 135-140.
[52] Ibid: 222.
[53] Ibid: 154-155.
[54] Johnson, Huron & Collister: 11-12.

than archaic words. 81% of the words on their "familiar" list were identified correctly, vs. 41% for words on their "archaic" list.[55]

Repetition

Immediate repetition facilitates word identification (a small effect—half of Johnson, Huron and Collister's subjects who misidentified a word on first hearing identified it correctly on the second, but almost two thirds as many identified the word correctly the first time and then misidentified it the second). Repetition after intervening material does not seem to facilitate identification, although their sample was small enough that this result is weak.[56] Interestingly, preceding a melismatic setting of a word by a syllabic setting of the same word does NOT increase identification of the melismatically set word. However, a syllabic setting after a melismatic setting DOES increase identification.[57]

In a slightly questionable portion of their experiment, Johnson et al. found that the second of two words is not more easily identifiable if it rhymes with the first. However, this is not how rhyme is actually used in lyrics, where rhyming words typically come at the ends of lines, often at rhythmically predictable times. The case of two adjacent words rhyming in sung text is a sufficiently rare occurrence that this result seems to have little bearing on real-world cases.[58] In this paper, I take the common sense view that the 2nd words of rhyming pairs in lyrics are likely to be *more* intelligible by virtue of that fact.[59]

---

55 Johnson, Huron & Collister, 7-8.
56 Ibid, 12-13.
57 Ibid, 13-14.
58 Ibid, 14-15.
59 A chain of reasoning supporting this idea appears in Chapter 4.

A further insight into the influence of repetition on intelligibility is offered in the 2015 paper by Huron and Nathaniel Condit-Shultz cited in the Introduction. It appears that

> listeners' interpretations of lyrics stabilize after an average of three hearings, beyond which they are no longer likely to improve their transcriptions. Rather, it seems that listeners simply become increasingly confident that their interpretation is correct, whether or not it actually is.[60]

Put another way, repetition beyond three times is not going to increase the intelligibility of a given word.

Harmonic Priming

An additional perspective comes from the work of cognitive psychologist E. Bigand et al.: they discuss the phenomenon of *harmonic priming.* In a 2001 paper, they asked participants to distinguish between two vowels at the end of chord progressions played back using sampled voice sounds. They found that the phoneme recognition was noticeably faster when the final chord was a tonic chord than when it was a subdominant chord.[61] A subsequent study by some of the same authors asked whether the difference in recognition between a lexically expected word at the end of a phrase vs. an unexpected word (semantic priming) would be affected by harmonic setting. They found that lexically expected words were recognized correctly more often than lexically unexpected words (a result I will discuss in greater detail in Chapter 5). Additionally, however, they found that, when the final chord of the phrase was a tonic chord, the recognition rate was

---

[60] Condit-Schultz and Huron: 482.
[61] E. Bigand, B. Tillmann, B. Poulin, D.A. D'Adamo, F. Madurell, "The effect of harmonic context on phoneme monitoring in vocal music," *Cognition* 81 (2001): B16.

higher, and the rate of recognition was significantly faster than when the final chord was a subdominant chord.[62] This is certainly suggestive; while it's not obvious how the result can be generalized to other musical contexts, a hint may be supplied by the fact that in both of these papers the tonic chord was considered the *related* or *referential condition,* contrasted with the *congruent but less related* or *referential condition* of the subdominant chord.

The "rate of recognition" in the results above refers to the commonly accepted notion in linguistics that there are finite cognitive resources available for the mental processing of language (or music), and that phonemes or words that take longer to process will thus be more difficult to understand, creating a bottleneck in the flow of processing. Condit-Schultz and Huron also note in discussing their results that "length in seconds [of their excerpts] did appear to be a significant predictor [of unintelligibility,] suggesting that longer excerpts may have suffered from memory limitations."[63] This also brings up the idea that limited cognitive resources may be an issue in the intelligibility of sung text, which is the subject of Chapter 3.

Environment

There remains the fact that classical music is usually sung in environments that are not conducive to intelligibility. The optimal reverberation time for understanding a speaker is about three-quarters of a second. The typical reverberation time of a concert hall or opera house—the "acoustically friendly environment" we are assuming—is about

[62] Bénédicte Poulin-Charronnat, Emmanuel Bigand, François Madurell, and Ronald Peereman, "Musical structure modulates semantic priming in vocal music," *Cognition* 94 (2005): B71-72.
[63] Condit-Schultz and Huron: 479.

twice that.[64] The longer reverberation times that give music a richness of texture also mask the source sound to some extent, a phenomenon known as *self-masking.* In the case of spoken language, self-masking is particularly hard on the fricatives and plosives, the consonants whose sonic intensity is among their primary identifying features,[65] although it is not clear how this effect would combine with the already mentioned de-intensification of these consonants in singing.

Chapters 1 and 2 reviewed the resources currently available from our own disciplines of composition and musicology towards an understanding of why classically trained singers are difficult to understand, and the magnitudes and natures of the various contributing effects' influences. These resources begin to illuminate the nature of and possible practical solutions to the problem, but only in patches.

Here's what we can bring forward to our growing understanding of how to maximize the intelligibility of set English text:

•above about F4, all vowels become /ɑ/, and even vowels sung below that pitch distort to some extent, following a set sequence;

•diphthongs are more easily identified than monophthongs;

•certain consonants, particularly voiced stops, are more vulnerable to mishearing than others;

•syllabic settings are more easily understood than melismatic ones;

•familiar words are more easily understood than archaic ones;

•repetition (up to three times) facilitates word recognition, unless the latter iterations of the word are melismatic settings;

---

[64] Vern O Knudsen, "The hearing of speech in auditoriums," *The Journal of the Acoustical Society of America* 1 (1929): 74-75.

[65] Anna K. Nábēlek, Tomasz R. Letowski, and Frances M. Tucker, "Reverberant overlap- and self-masking in consonant identification," *The Journal of the Acoustical Society of America* 86 (1989): 1264.

•and the expectedness or unexpectedness of harmonic settings can influence the recognition of sung words.

This is a small treasure trove of insight to add to the received wisdom of composition teachers, as exemplified by Thomson, Barker and Rosenhaus. But it is far from sufficient to our purposes. What other resources can we bring to bear on the problem?

Chapters 3, 4 and 5 will draw on sources from other disciplines—cognitive science, phonetics and linguistics—to round out the picture, giving the composer who wishes to maximize the intelligibility of an English text setting a more complete understanding of both the obstacles involved and the possible means of surmounting or avoiding them.

Chapter 3: Concepts from Cognitive Science

Could part of the reason for the unintelligibility of sung text simply be that it is more difficult to understand words when one is also hearing music? Language and music share certain properties, notably the presence of syntax and the use of working memory. If the cognitive processing of both musical and linguistic syntax uses the same resources, then text settings that were syntactically complex in both ways might prove particularly difficult to understand. Similarly, if the amount of material that can be stored in working memory has limits, then we ought to be able to discern a maximum amount of information that can be apprehended per unit of time (information density)—and if musical and linguistic working memory are not completely independent, then it will be the total information density, musical and linguistic combined, that matters, rather than the density of one or the other. As it happens, cognitive scientists have indeed explored these questions, and there are conclusions we can draw from their results that could be useful to composers and analysts.

Cognitive Resources

As mentioned in Chapter 2, it is a commonly accepted idea in linguistics that there are finite cognitive resources available for the processing of language, and that portions of linguistic tasks that require more resources will produce a bottleneck in the processing flow. Eye-tracking studies, in which the eye motions of subjects are analyzed while they are reading a text to see which words they unconsciously linger on, are a staple of the literature; more recently, fMRI studies of subjects listening to recorded language (to be discussed below) have obtained similar results, using measures of blood flow to brain regions associated with hearing as the audio equivalent of a lingering eye. The central insight here is that certain words take more resources to process than others; they are more difficult to understand, whether that difficulty comes in recognizing or comprehending the word or in sorting it into its appropriate syntactic role.

Although fewer studies have investigated this, it seems logical that musical processing is subject to similar constraints. The question for our purposes, though, is whether language and music *share cognitive resources*, such that the processing of music interferes with the processing of language. There is evidence that music and language share resources, and there is evidence that they have separate resources. It seems most likely that there are some areas that are shared and some that are separate.

Evidence for shared resources

In the literature review for a paper describing the results of a 2010 set of fMRI experiments—the first specifically comparing the perception of song, speech and vocalise—neuroscientist Daniele Schön et al. ran through a bewildering array of prior contradictory results: some studies indicated that semantic and melodic or harmonic processing were independent. Others pointed to the interaction of musical and linguistic syntax and semantics, and the interaction of phonological, lexico-semantic and syntactic processing with harmonic processing.[66]

---

[66] Daniele Schön, Reyna Gordon, Aurélie Campagne, Cyrille Magne, Corine Astésano, Jean-Luc Anton, Mireille Besson, "Similar cerebral networks in language, music and song perception," *NeuroImage* 51 (2010): 459-460.

Schön and his colleagues' own study showed that a "common network of brain regions" is involved in the perception of all three, and that in certain areas of the brain perception of music and language interact with each other: "the linguistic and musical dimensions of sung words were not processed independently." (Specifically, the interaction was between the lexical/phonological aspects of language processing and the melodic aspect of musical processing.)[67] Interestingly, song and speech showed greater similarities in the activation of brain areas than did song and vocalise; but all of the brain areas under investigation were involved in all three types of perception. The differences were of degree rather than kind.[68]

A 2009 study by psychologist Régine Kolinsky et al. advanced a fascinating proposal as to the nature of the interaction. They found that (in singing) consonants are processed more independently from melody than are vowels.[69] Since learning and developmental research imply strongly that "vowels and consonants subtend different linguistic functions, with consonants being more tied to word identification, while vowels essentially contribute to grammar and to prosody,"[70] it appears that vowels and melodic intervals may share a syntactic/grammatical role in the apprehension of singing, whereas consonants' function is more lexical.[71]

---

[67] Ibid: 455-456.
[68] Ibid: 454-455.
[69] Régine Kolinsky, Pascale Lidji, Isabelle Peretz, Mireille Besson & José Morais, "Processing interactions between phonology and melody: Vowels sing but consonants speak," *Cognition* 112 (2009): 15.
[70] Ibid: 2.
[71] Ibid: 17.

Evidence for separate resources

Lacking any way to discern directly what brain regions are functioning during musical and verbal tasks, in the 1990s music education researcher William Berz employed some ingenious deduction. Referencing experiments in which background music was played while subjects performed verbal tasks, he concluded that there must be a working memory for music different than that for speech: "If there was a singular acoustic store, unattended instrumental music would cause the same disruptions on verbal performance as would unattended speech or unattended vocal music; this was shown not to be the case."[72]

Technological advances have confirmed Berz's conclusion. Recently, a groundbreaking study demonstrated the physical existence of a part of the auditory cortex devoted solely to the processing of music. Participants were played recordings of multiple sound clips of various sorts, some speech, some noise, some music. When the clips of the sounds that had engendered similar fMRI responses were compared, the musical clips were found to have excited the same neural pathway, a distinct pathway from that excited by speech.[73]

Reconciling varying results

Psychologist Patel Aniruddh hypothesized in 1998 that "linguistic and musical syntactic processing rely on distinct cognitive operations, but structural integration in

---

[72] William L. Berz, "Working Memory in Music: A Theoretical Model," *Music Perception* 12 (1995): 361.
[73] Sam Norman-Haignere, Nancy G. Kanwisher and Josh H. McDermott, "Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition," *Neuron* 88 (2015): 1285.

both domains relies on a common pool of neural resources."[74] In a 2003 paper, he refined that hypothesis to propose a structure in which language and music share a module responsible for syntax (the relation of events to other events at different times), but have their own modules for knowledge of their separate domains.[75] Linguist Ray Jackendoff, taking a more *a priori* approach to the question in 2009, argued that most of the characteristics that language and music share do "not indicate a particularly close relation that makes them distinct from other cognitive domains. Many shared characteristics are domain-general, for instance recursion, the use of memory, and the need for learning and for a social context."[76]

It is not clear whether Berz's "musical working memory" or Patel's "domain-specific module" have any correlation with the music-specific neural pathways discovered by Sam Norman-Haignere and his neuroscientist colleagues. For the purposes of this paper, I will adopt the viewpoint that music and language processing share some, though not all, cognitive resources, and thus are in a position to interfere with each other under certain circumstances, contributing to the difficulty of understanding sung language. Patel's notion that syntax is at the core of the cognitive overlap seems particularly fruitful; perhaps syntactically complex language combined with syntactically complex music would pose a particular challenge to intelligibility.

---

[74] Aniruddh D. Patel, "Syntactic processing in language and music: different cognitive operations, similar neural resources?" *Music Perception* 16 (1998): 29.
[75] Aniruddh D. Patel, "Language, music, syntax and the brain," *Nature Neuroscience* 6 (2003): 676.
[76] Ray Jackendoff, "Parallels and nonparallels between language and music" *Music Perception* 26 (2009): 203.

Working memory

If there is a musical working memory separate from linguistic working memory, how do they function when they are both engaged in listening to the same thing—when listening to sung text? Conclusions about working memory (also known as short-term memory, or STM) in music tend to be generalizations from broader studies of memory, including linguistic memory. Could any of these conclusions prove useful in the study of intelligibility?

There have been two theories about how working memory reaches its limits. One is that working memory has a finite capacity for items of whatever sort, traditionally called "chunks;" the other is that working memory is limited simply by the amount of time elapsed. (Working memory can also be limited by attention, but as we are considering the case of an engaged, attentive listener, we will disregard this aspect.) In the 1950s, it was determined that seven random words was the average limit for short term memory,[77] a result that made its way into pop culture. This was translated by subsequent research into a limit of three to five chunks. How much information is contained in a chunk can vary, as the size of a chunk becomes larger the more familiar one is with the material being remembered.[78] The limit in time is startlingly short—about two seconds. However, it appears that rather than simply a decay of memory over time, the effect is due to the interference of intervening stimuli. (Going over information repeatedly keeps it in working memory, a process called *rehearsal* in this branch of

[77] Nelson Cowan, Candice C. Morey, Zhijian Chen, Amanda L. Gilchrist and J. Scott Saults, "Theory and Measurement of Working Memory Capacity Limits,." *The Psychology of Learning and Motivation* 49 (2008): 56.
[78] Ibid: 59-60.

cognitive science.)[79] In a 2005 paper, molecular biologist Zhijian Chen and psychologist Nelson Cowan showed that the two models are complementary—that neither can explain all test results alone, but together they cover all circumstances.[80]

Summing up: there is a separate neural pathway for the perception of musical sound; there are separate working memory stores for musical and linguistic information; the processing of musical and linguistic information takes place in a shared network of brain areas, with vowels and melodic intervals interacting more than consonants; and working memory of whatever kind has capacity limits based on both elapsed time and on number of chunks. What might we deduce from this?

The information in many musical contexts can be analyzed in terms of chunks per unit of time; as I have not seen this concept anywhere in the linguistics literature, I will call this *chunk density*.[81] One conclusion might be that if either linguistic or musical chunk density exceeds the limit of working memory, or if the combined chunk density exceeds the limit of the shared network to process the input, there will be a bottleneck and comprehension will fail, whether of the words, the music, or both. A composer wanting to create a musically dense but intelligible passage, for instance, might take care to include no more than 3-5 musical chunks in a given 2-second period by making sure that the material has been presented previously with sufficient repetition that it is indeed perceivable as chunks rather than as the chunks' constituent components.

---

[79] Ibid: 77-79.
[80] Zhijian Chen and Nelson Cowan, "Chunk Limits and Length Limits in Immediate Recall: A Reconciliation," *Journal of Experimental Psychology* 31 (2005): 1247.
[81] The term "chunk density" does have a different meaning in linguistics, where it refers to the prevalence of multi-word chunks in a language corpus.

Another conclusion might be that if the retention of information longer than two seconds is desirable, as for instance in the setting of polysyllabic words at slow tempi, where the listener will need to still remember the beginning of the word by the time s/he hears the end in order to understand it, that a composer should take care either to repeat (rehearse) the information and/or to keep musical events such as accompaniment sparse in order to reduce the effect of the interference of other stimuli.

This brief excursion into cognitive science has yielded some concrete propositions to add to our toolkit for maximizing intelligibility of sung text. There are still other possibilities inherent in the notion that vowels and intervals influence the perception of prosody and grammar, whereas consonants' role is in word identification. In order to grapple with the implications, however, we will need to go into greater detail about how we perceive both consonants and vowels, and how we identify both words and grammatical structures. These are the subjects of the next two chapters.

Chapter 4: Phonemes on Parade

Some words are more vulnerable to mishearing than others, making their intelligibility more problematic. What is it about these words that makes them so slippery? An investigation into phonology and linguistics reveals that certain phonemes are more vulnerable to misinterpretation than others, especially when sung. These phonemes are most likely to disrupt the workings of the word recognition process, and should be treated carefully by the composer trying to include them in an intelligible text setting. Moreover, the context of the word can influence the perception of its phonemes!

As mentioned previously, both consonants and vowels can be distorted from their spoken forms in singing. A brief exploration into the mechanisms by which we understand spoken words will help illustrate some of the specific challenges this phenomenon provides in understanding sung words, which will in turn contribute to our knowledge of how to surmount them. To first understand how word recognition works when the words are spoken and heard clearly, we will examine the Cohort Model of spoken word recognition. We will then review some phonological studies for specific information about which phonemes are most difficult to process, and in which contexts.

The Cohort Model

In the 1980s, cognitive scientist William D. Marslen-Wilson introduced the Cohort Model of spoken word recognition, which was then refined by himself and others. The essence of the Cohort Model is that when we hear the initial phonemes of a word, all of the possible matches for those phonemes among the words we know are neurologically activated; as we hear more of the word, the number of possible matches is reduced, until at some point (not necessarily the end of the word) we have eliminated all possibilities except one. So, for instance, when we hear the word "banana," at the end of the first syllable (/bə/) "banana" has been activated, but so have "ballistic, "banal," and "bedraggled." After hearing the first /n/ the field has been narrowed down to "banana" and "banal;" as soon as we hear the vowel of the second syllable, however, "banana" emerges as the only match.[82] Thus the sound made by the second "a" is the *recognition point* of the word "banana." (The collection of words that match a given string of auditory input is the "cohort" of the model's title.)

One of the principal refinements to the model was the acknowledgement that context also plays a role in the recognition of words; perception is not just a bottom-up affair. If we hear the word in an aria about fruit, "banana" will have an edge over "ballistic," even after the first syllable; if the aria is about weapons systems, the reverse will be true. However, auditory input overrules context:

> If the sensory input clearly differentiates one candidate from all others, then that is the candidate that will emerge from the perceptual process, irrespective of the degree of contextual anomaly. If contextual variables clearly indicate a given candidate, it will nonetheless not emerge as the choice of the system unless it also fits the bottom-up input.[83]

If the context does not make the word clear, though, the composer should be aware of the recognition point of the word. If s/he asks the singer to hold out, say, the first syllable of a word whose recognition point is in the second syllable (i.e. *different* vs.

[82] Marcus Taft and Gail Hambly, "Exploring the cohort model of spoken word recognition," *Cognition* 22 (1986): 260.
[83] William D. Marslen-Wilson, "Functional parallelism in spoken word recognition," *Cognition* 25 (1987): 97.

difficult), s/he is asking the listener to suspend recognition of the word until that recognition point is reached, putting a strain on the perceptual process. Moreover, as we saw in the last chapter, if the held note is longer than two seconds and accompanied by more than three to five "chunks" of musical information, the listener is in danger of forgetting the first syllable of the word before the recognition point is even reached!

The influence of context

What about cases when the sensory input does *not* clearly differentiate one candidate from all others—when what one hears is ambiguous? Cognitive scientists Marcus Taft and Gail Hambly tested perception of non-word phoneme strings, and found that

> subjects occasionally confused the nonwords with real words, and this was more likely to happen with the shorter nonwords, since these had a greater proportion of phonemes in common with real words than did the longer nonwords (e.g., MEP being similar to MAP and MET).[84]

Extrapolating from these results, we can surmise that in cases in which the auditory input is not clear—often the case with sung language, with its modified vowels and consonants, as discussed in Chapter 2—the perceptual process will either have some result imposed top-down by context (as long as the auditory input does not preclude it), result in an incorrect match (a "monde-green")[85], or result in no match at all.

---

[84] Taft and Hambly: 266.
[85] A "mondegreen" is a substitution of words of a song as a result of mishearing. The term was coined by Sylvia Wright in the 1950s, and comes from her substitution of "Lady Mondegreen" for "laid him on the green" in a Scottish ballad (Dictionary.com, accessed 4/29/2016).

There are other implications of the Cohort Model and its supporting evidence potentially relevant to the understanding of sung language. Words that are more frequently used come to the top of the cohort faster than words that are rarer, in much the same way that words reinforced by context have an edge (recall our discussion of "banana" vs. "ballistic"). High-frequency words that have the same recognition point as lower-frequency words are nonetheless recognized faster.[86] This nicely confirms Collister and Huron's findings about the difference in intelligibility between common and archaic words.[87] Furthermore, strings of phonemes do not have to actually be words to activate neurolexical entries: in the phrase "new disturbance," for instance, the word "nudist" will be activated.[88] This immediately suggests that a composer setting text should be able to take steps to minimize this kind of spurious activation by creating a discontinuity, whether rhythmic or melodic, between the words "new" and "disturbance."

Can we be specific about what types of words are more likely to be misperceived? As a matter of fact, we can. Words that begin with voiced stops or unvoiced fricatives are vulnerable, as are words beginning with consonants and followed by /i/ or another front vowel—particularly words whose meanings are not reinforced by context. We can derive these interesting guidelines by examining results from phonology through the lens of what we know from Chapter 2 about how singing affects the production and perception of various phonemes.

---

[86] Taft and Hambly: 270-271.
[87] The same is true of high-frequency grammatical structures, the implications of which I will discuss in Chapter 5. (Altmann: 96.)
[88] Gerry T. M. Altmann, *The Ascent of Babel,* Oxford University Press (New York: 1997): 82.

Ambiguity of consonants

In search of ways to make spoken communication over noisy channels more reliable, in the 1950s psychologists George A. Miller and Patricia E. Nicely did an exhaustive study of the perception of consonants through noise (and also through high-pass and low-pass filters). Their results indicated that different articulatory features, which linguists use to divide phonemes into families, are affected differently by noise.[89] The features they identified were nasality, voicing, affrication, duration, and place of articulation. Of these, nasality and voicing were the least affected, and place of articulation was the most affected—they characterized it as "severely" so. [90]

What this means is that consonants which are distinguished from each other only by their place of articulation (front, middle or back of the oral cavity), such as the voiced or voiceless stops, are the most vulnerable to being disrupted by interference. In other words, /b/, /d/, and /g/ (the voiced stops) are easily confused with each other, and /p/, /t/, and /k/ (the unvoiced stops) are likewise easily mistaken for each other. Because voicing is among the least affected features, it is unlikely that voiced and voiceless consonants will be mistaken for each other. /b/ is more likely to be heard as /d/ or /g/ than as /p/; /p/ is more likely to be heard as /t/ or /k/ than as /b/.

---

[89] The analogy between the effect of noise, such as static, and musical sound on the transmission of language is tremendously suggestive, but to my knowledge there have been no studies that directly compare them. Hence, I will not make direct use of the many specific results from linguistics and information theory that discuss intelligibility through noise, noting only that it is likely that there is some similarity of effect.

[90] George A. Miller and Patricia E. Nicely, "An Analysis of Perceptual Confusions Among Some English Consonants," *The Journal of the Acoustical Society of America* 27 (1955): 338.

You may recall from the discussion of Huron and Collister's work in Chapter 2 that the voiced stops were the category with the largest number of errors in their tests. Miller and Nicely's work offers a reason for this:

> For the voiced stops |bdg| the most important acoustic clue to position seems to be in the initial portion of the second formant of the vowel |a| that follows; if this formant frequency rises initially, it is a |b|, but if it falls it is |d| or |g|. *Since the vowel formant is relatively audible,* [my emphasis], the front |b| is easily distinguished from the middle |d| and the back |g|. The latter two positions are much harder to distinguish…[91]

As also discussed in Chapter 2, the formant transition—the "initial portion of the vowel formant frequency"—is often shorter than in speech. On high notes, the vowel formant is often absent altogether, replaced by the fundamental of the note being sung. Thus the cues that help to distinguish place of articulation in consonants—that help to distinguish between members of the family of voiced stops—are either lessened or obscured in singing.

How does this effect interact with the Cohort Model? Logically, there are two possibilities; either the consonant will be misheard as one of its family members, activating the wrong cohort, or it will be heard as ambiguous, not activating *any* cohort until more information from later in the word comes in and fills it in retroactively. This suggests that confusion of initial consonants is most likely to derail intelligibility completely.

Here we have the makings of one possible rule of thumb towards maximizing intelligibility of sung text: be on the lookout for words using voiced stops, especially those beginning with voiced stops. Avoid setting them on high notes unless the level of

---

[91] Ibid: 347. Their results were similar for voiced stops preceded by vowels.

possible ambiguity is low, whether because the combined cohort of possibilities for the word given the possible mishearings of the consonant is small or because context narrows down the possibilities. Since unvoiced fricatives (/f/, /s/, /θ/ and /ʃ/) are the other category of consonants identified by di Carlo as being particularly vulnerable to mishearing during singing, similar care can be taken with them; high notes and loud notes are the settings most likely to engender confusion here.

Effects of vowels on consonant intelligibility

In another study of the perceptual features of consonants heard with and without accompanying noise, psychologists Marilyn D. Wang and Robert C. Bilger discussed the effects of accompanying vowels on the identification of consonants. Both with and without noise, they found that "consonants were better identified when the accompanying vowel was /u/ rather than /ɑ/." This was true for both CV (consonant/vowel) and VC (vowel/consonant) combinations. However, "in CV syllable sets, consonants followed by /i/ were the most difficult to identify, whereas in the VC syllable sets, consonants preceded by /i/ were the most easily identified."[92]

Wang and Bilger's syllable sets used only the three vowels mentioned above, along with 16 consonants; unfortunately, there are no equivalent data sets for other vowels in combination with consonants. However, those three vowels were chosen because they represent the extremes of articulatory positioning of the tongue and vocal

---

[92] Marilyn D. Wang and Robert C. Bilger, "Consonant confusions in noise: a study of perceptual features," *The Journal of the Acoustical Society of America* 54 (1973): 1251-1252.

tract. English vowels are traditionally represented on a trapezoidal diagram representing the oral cavity seen from the side with the lips on the left. The position on the trapezoid represents the articulatory position of the vowel; from left to right represents the front of the oral cavity to the back, and from top to bottom represents the spectrum from closed vowel to open vowel. Of Wang and Bilger's three vowels, /i/ is front-close, /ɑ/ is back-open, and /u/ is back-close: the upper-left, lower-right and upper-right corners, respectively (see Figure 1 below).

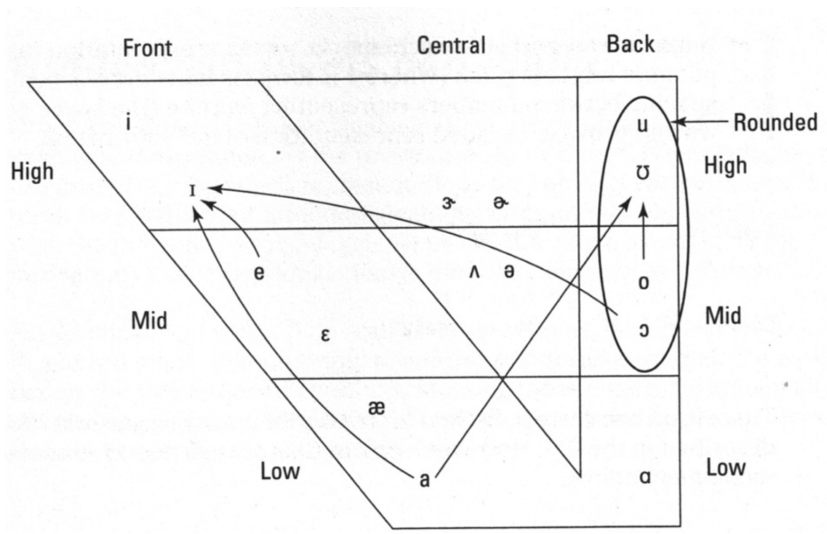


Figure 1. Vowel Quadrilateral. Source: William F. Katz, *Phonetics for Dummies* (Hoboken: John Wiley & Sons, 2013) 78 (fig. 5-3).

This means that we should be able to interpolate from Wang and Bilger's results. If consonants are more easily identified when accompanied by /u/ than by /ɑ/, then it seems likely that consonants accompanied by back vowels will be more easily identified as those vowels climb the back of the trapezoid from /ɑ/ to /u/. If consonants preceded by /i/ are easier to identify than consonants preceded by /ɑ/ or /u/, then we can surmise that identification for those latter consonants will become easier as the preceding vowels move from back to front and/or from open to close. If consonants followed by /i/ are harder to identify than consonants followed by /ɑ/ or /u/, then we can surmise that identification for those latter consonants will become harder as the preceding vowels move from front to back and/or from close to open.

We need to remember, of course, that these results are for spoken language, and our understanding of them needs to be modified to be useful for the task of maximizing the intelligibility of sung text. (As pitch shifts upward and the fundamental of the sung note overtakes the first formant of the sung vowel, every vowel begins to behave more like /ɑ/.) But it seems clear that these results, along with a working knowledge of the placement of vowels, will enable the vocal composer to develop a sense of which consonants in which words are most at risk of being misheard. Of particular interest is the information that words beginning with consonants and followed by /i/ or another vowel near it on the diagram are more likely than most words to be perceived as beginning with a different consonant, thus activating the wrong cohort and subject to being misunderstood.

Ambiguity of vowels

We have already noted that the perception of vowels can be drastically affected by pitch, and that misidentification can result, contributing to unintelligibility. But is there any way to minimize this effect? It turns out that there is.

Biophysicists John Smith and Joe Wolfe note that vowels have natural resonance bands, and that vowels will be most identifiable when produced on notes within their natural range of frequencies. (At other frequencies, singers will want to tune the resonance of the vowel to maintain vocal power and homogeneity of tone.) They categorize vowels by their place on the open/close spectrum (see Figure 1) and provide the following values: closed 250–400 Hz, close–mid 400–550 Hz, open–mid 550–750 Hz, and open 750–1000 Hz.[93] Thus the vowels /i/ and /u/ will be most readily identifiable between the notes B3 and F#4, the vowel /I/ between F#4 and C#5, /ɛ/ between C#5 and F#5, and /ɑ/ between F#5 and B5. A composer wishing to maximize intelligibility could take advantage of this phenomenon.[94]

Vowel timing

Cognitive psychologists Winifred Strange, James L. Jenkins and Thomas L. Johnson found that, if you cut out the middle of the vowel of a recorded consonant-vowel-consonant (CVC) combination, and splice the remaining portions together, it is still possible to understand the syllable—the "dynamic transitions" into and out of the vowel carry sufficient information to recognize the vowel.[95] This would seem to indicate that the length of the vowel does not matter. However, they go on to note that

> Vowels, as gestures, are differentiated by their timing with respect to adjacent segments and syllables, as well as by the positioning of the tongue during the relatively sustained vocalic portion of the syllable. The perceiver must identify the intended vowels on the basis of information in the acoustic pattern about the *timing* of the gestures as well as the vocal tract state attained.[96]

---

[93] John Smith and Joe Wolfe, "Wagner's music is even better than it sounds: implications of vowel-pitch matching for intelligibility and ease of singing," *The Second International Conference on Music Communication Science* (2009): 92-95.
[94] As it is difficult to imagine a setting that adheres strictly to this rule—what about diphthongs, for one thing?—I will not be including this phenomenon in the analyses that end this paper.
[95] Winifred Strange, James L. Jenkins and Thomas L. Johnson, "Dynamic specification of coarticulated vowels," *The Journal of the Acoustical Society of America* 74 (1983): 695. Not only that, the dynamic spectra of transitions into and out of vowels are actually *better* than the steady-state nuclei of the vowels at conveying the identity of the vowels. (Winifred Strange, ""Dynamic specification of coarticulated vowels spoken in sentence context," *The Journal of the Acoustical Society of America* 85 (1989): 2146.)
[96] Ibid: 704.

This indicates that vocal music sung in a rhythm approximating that of speech, in which the timing of the vowels "with respect to their adjacent segments and syllables," will be easier to understand than music in which the relative timing of vowels is greatly different than that of speech, and gives an after-the-fact rationale for the use of speechlike recitative in a relatively low vocal range to put across plot points in centuries' worth of opera.

The Cohort Model goes quite a way toward explaining why the 2nd members of rhyming pairs at the ends of lyric lines should be more intelligible. If the placement of the rhyme is predictable (as it often is), so that the listener can predict when the rhyming word is coming, the cohort for that word will be drastically reduced compared to non-rhyming words; indeed, once any initial consonant phonemes have been determined, the processing cost for that word is zero.

The workings of the Cohort Model, along with extrapolations from some specific results from phonology and acoustics, are beginning to give us some rules of thumb for understanding which words of a text are likely to be most difficult to process, and thus should be set with extra care. However, just as the difficulty of processing—the recognition—of phonemes is influenced by neighboring phonemes, so it turns out is the difficulty of processing of words influenced by neighboring words. The Cohort Model can (metaphorically, at least) be extended to the level of the sentence: grammatical structures have recognition points similar to those of words. In Chapter 5 we will explore the implications of this for the understanding of sung text.

Chapter 5: At The Level of the Sentence

An effect very like the Cohort Model influences the understanding of language at the level of the sentence. As we hear language, we are engaged in a continuous process of prediction, using what we have heard so far to guess what comes next, attempting to stay even with and even get ahead of the inrushing flow of words. Unexpected words slow down this process, potentially creating a bottleneck in comprehension. Here again is a phenomenon that the composer who wants to maximize intelligibility can take into account when setting text: words that are unexpected in this regard are more likely to fall prey to all the forces already discussed that work against comprehension. Once identified, they can be set with extra care (in combination with the other perspectives already discussed). In order to understand which words are unexpected, though, we need to examine some results from the field of psycholinguistics—the study of the interaction between language and psychological processes.

Word Frequency

One of the earliest results in the field was the finding that words that occur more frequently in the language are easier to identify than less common words. In a 1957 paper, psychologists Mark R. Rosenzweig and Leo Postman discussed the results of an investigation of intelligibility through masking noise similar to that of Miller and Nicely (discussed in Chapter 4), but this time comparing words by their frequency of occurrence in the language. They found that “The most frequent words could be perceived, on the average, through about 12 db more noise than could the least frequent words."[97] (Again, the equivalence of masking noise to accompanying musical information is unprovable, but suggestive.)

However, there was a wrinkle: longer words were well known to be often more intelligible than shorter ones, despite their being usually less common in the language. There seemed to be two different influences on intelligibility through noise: word frequency and word complexity (longer words are more complex), both independent of the word's phonetic composition. Also in 1957, Rosenszweig and Postman's colleague Davis Howes coined the term "effective relative frequency" to describe the influence of a word's length on its intelligibility:

> The effective relative frequency of a word thus depends on the sum of the frequencies of all words that cannot be discriminated from it with respect to length. As this sum is greatest for short words, the effective relative frequency of a short word is lower than that of a longer word of the same…frequency.[98]

This is the same basic idea as the Cohort Model: the ease of discrimination of a word depends on its competition from similar words. Longer words have fewer competitors than short words, so they are less prone to mishearing. This is a useful thing for a composer to know, and not part of the received wisdom represented by the Thomson and Barker books examined in Chapter 1.

It was shortly thereafter noted by psychologists Herbert Rubenstein and Irwin Pollack that the relative frequencies of words according to word lists (even adjusted for

---

[97] Mark R. Rosenzweig and Leo Postman, "Intelligibility as a function of frequency of usage," *Journal of Experimental Psychology* 54 (1957): 415.

[98]Davis Howes, "On the Relation between the Intelligibility and Frequency of Occurrence of English Word, "*The Journal of the Acoustical Society of America* 29 (1957): 303-304.

length) did not tell the whole story: it was easier to discriminate words in context—in sentences—than it was in isolation, and it was easier to discriminate even isolated words in a list when the words were all from a known category than when the list was mixed.[99] (This finding might be directly of use to a composer setting non-narrative poetry or the like.) However, their most striking result was that "small changes in predictability at the lower end of the scale were associated with large changes in intelligibility." They speculated that the relationship between predictability and intelligibility might turn out to be described by a logarithmic relationship.[100] In the 1990s the new field of computational psycholinguistics confirmed this and put the correlation on a sound mathematical basis with the introduction of the concept of *surprisal.*

Predictability: Surprisal and Entropy[101]

One of the primary enterprises of computational psycholinguistics has been to put human language comprehension on a mathematical basis by creating computer models that approximate the process. Modern computational power is used to apply artificial (simplified) grammars to large data sets of (usually written) language in order to make predictions about how language is perceived. The tools that are created to apply these grammars to these data sets are called "complexity metrics": they quantify how difficult it is to perceive a given linguistic expression. Specifically, these are "information-

---

[99]Herbert Rubenstein and Irwin Pollack, "Word Predictability and Intelligibility," *Journal of Verbal Learning and Verbal Behavior* 2 (1963): 147.
[100]Ibid, 148.
[101]I am indebted for much of the following discussion to John T. Hale, who graciously let me read a then-unpublished draft of a review article he was writing on the subject, and answered several follow-up questions by email.

theoretical complexity metrics": they apply concepts from information theory. (I will discuss which concepts shortly). The endgame of this process is to compare the predictions to observed measurements of linguistic processing difficulty such as reading time in eye-tracking studies or neural signals in fMRI experiments; if the predictions match up well with the experimental data, then the model that produced the predictions is coming close to describing the way humans actually process language.[102] (It is worth noting here that these models have, in fact, developed into quite effective tools; much of the recent impressive advance in computer recognition of spoken language and predictive text is attributable in part to this type of work.)

As daunting as a term like "information-theoretical complexity metric" is to a non-specialist, the idea behind the term is not difficult to follow. Each word in a sentence has a certain probability of occurring. This probability is based not only on its frequency in the language but on the words preceding it. Thus, even if "letter" and "brother" have similar frequencies in the English language, one has a much greater probability than the other of being the final word in the sentence "I went to the post office to mail my ______."

This concept has been formalized in the term *surprisal,* which is the logarithm of the reciprocal of the probability of a given word being the next word in a sentence—a low-probability word will have a high surprisal, and a high-probability word will have a low surprisal. This can be calculated based on an arbitrary number of words to the left of the word in question (an "n-gram"), or in the context of the sentence as a whole.[103]

---

[102] John T. Hale, "Information-theoretical complexity metrics," Unpublished draft of review article, 2015: 1-2.
[103]Ibid, 2-3.

There are two types of surprisal: lexical and syntactic. In the letter vs. brother example above, it is the lexical surprisal of the two words that are different. The syntactic surprisal of a word has to do with how hard it is to create a syntactic structure that integrates the word into what has come before. For instance, if we already have the words "The girl liked those," the word "cakes" in the next slot would have a lower syntactic surprisal than the word "who." The syntactic and lexical surprisal of a word together comprise its overall surprisal.[104] One study found that syntactic surprisal was a much larger effect than lexical surprisal, although they noted that this might be due to their use of a "relatively small" language corpus.[105]

A different but related model of how the unexpectedness of words in sentences plays out is based on the concept of *entropy.* Entropy—the amount of disorder—is a familiar idea from thermodynamics, applied to information theory as a measure of uncertainty by Claude Shannon in the 1940s. The idea in psycholinguistics, by way of the information-theoretical version, is that each new word in a sentence reduces uncertainty about what the rest of the sentence will be to some extent.[106] The amount of this reduction in uncertainty is proportional to the amount of difficulty a "sentence understander" will have processing the sentence at that word. The greater the reduction in uncertainty, the faster the processing.[107] (A different formulation considers the amount of entropy at each word, rather than the change in entropy.)[108] As an example: the fourth

---

[104] Brian Roark, Asaf Bachrach, Carlos Cardenas, and Christophe Pallier, "Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing," In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*: 327. Association for Computational Linguistics, 2009.
[105] Ibid: 331.
[106] John Hale, "Uncertainty about the rest of the sentence," *Cognitive Science* 30 (2006): 645.
[107] Ibid: 650.
[108] Roark, Bachrach, Cardenas and Pallier: 331.

word of the sentence “The dog barked loudly” brings the entropy very low, whereas the fourth word of the sentence beginning “The dog barked and…” brings the entropy much higher. For most sentences, the results of surprisal or entropy analysis will be very similar. (Results may differ more in the case of sentences with *long-range dependencies* such as relative clauses; further discussion of these cases follows shortly.)[109]

Some of the computer programs (“parsers”) that do these analyses, and the corpuses of written language that they consult, are available to the public, although they generally do not have user-friendly interfaces. A composer who wishes to investigate the surprisal and/or entropy scores of the words in a text will be able to use these tools with a little work; alternatively, the concepts themselves might be enough for an informal analysis of a text.[110] This is the approach I will take in the analyses later in this paper.

Auditory language comprehension and predictability

We have learned that words with high surprisal scores, or words that increase entropy, are likely to be more difficult to process than other parts of a sentence. Most of the above research, however, was originally applied to the rate of comprehension of written words. Does it apply equally to spoken, and thus (presumably) to sung language as well? The answer is yes. In a 2015 experiment, John T. Hale et al used fMRI to correlate listeners’ hearing of an audiobook with blood flow to the listeners’ language-related brain regions, and compared the results with the predictions made by several levels of linguistic modeling. They concluded that humans use the same type of linguistic

[109] Roger Levy, “Expectation-based syntactic comprehension,” *Cognition* 106 (2007): 35.
[110] The Roark Parser, for instance, is available at https://github.com/roarkbr/incremental-top-down-parser, including a version with the Penn WSJ Treebank corpus.

structures (as modeled by the theories discussed above) to process a spoken narrative as a written one.[111]

Moreover, “context information also changes the brain’s scalp-recorded electrophysiological response to words.”[112] An analysis of studies of the N400 ERP response (one type of these responses) by psychologist Kara D. Federmeier showed “a strong, inverse correlation with the predictability of the eliciting word within a given context” in comprehension of both the spoken and written word.[113] (Measures of predictability varied in the studies Federmeier cited, but used similar rubrics to those discussed above.) This means that, whatever the mechanism, the brain’s responses to language on the level of predictability are the same whether the words are written or spoken, supporting the conclusion reached by Hale and his colleagues and relating it specifically to the concept of predictability (or *expectancy*), the quality quantified by surprisal and entropy.

Surprisal, in particular, has been found to have a correlation with a characteristic of spoken language: word duration. Computational linguist Vera Demberg and her colleagues found that syntactic surprisal predicted changes in spoken, conversational word length better than a state-of-the-art text-to-speech system, word frequencies, or probabilities based on trigrams (the frequencies of words based on the two preceding words).[114] Words with higher surprisal are generally spoken more slowly and clearly than

---

[111] John T. Hale, David E. Lutz, Wen-Ming Luh, and Jonathan R. Brennan, "Modeling fMRI time courses with linguistic structure at various grain sizes," *Proceedings of CMCL 2015*: 95.
[112] Kara D. Federmeier, "Thinking ahead: The role and roots of prediction in language comprehension," *Psychophysiology* 44 (2007): 491.
[113] Ibid: 492.
[114] Vera Demberg, Asad B. Sayeed, Philip J. Gorinski, and Nikolaos Engonopolous, “Syntactic surprisal affects spoken word duration in conversational contexts,” *Proceedings of Sixth ISCA Workshop on Speech Synthesis*, 1.

canonical word duration and word frequency would lead one to expect.[115] Because words with high surprisal carry more information (in the information-theoretical sense), this result supports a notion called the Uniform Information Density hypothesis: that speakers try to distribute information uniformly across their utterances. While the effect is not huge (surprisal scores of 2.179 and 16.277 for the word "thing" in two different contexts imply a duration difference of about a tenth of a second), it is certainly audible.[116] This information might prove useful to a composer in choosing a rhythmic setting for a high-surprisal word, for instance.

Limitations and competing theories

Is the surprisal/entropy description of expectancy the last word on what makes certain words more difficult to understand than others? Actually, there is another, competing set of theories about the causes of linguistic processing bottlenecks. In these theories, memory is the scarce resource, and words that require reference to other, previously encountered words in a sentence ("dependencies"), whether syntactically or lexically, are the words that create the bottlenecks.[117] In one of them, Dependency Locality Theory (DLT), there are two types of costs in the process of comprehending a sentence: the cost of establishing the dependencies between current words and earlier parts of the sentence; and the cost of maintaining representations of incomplete

---

[115] Ibid: 5.
[116] Ibid: 5.
[117] Daniel Grodner and Edward Gibson, "Consequences of the Serial Nature of Linguistic Input for Sentenial [sic] Complexity," *Cognitive Science* 29 (2005): 261.

dependencies until they are completed.[118] Sentences with complicated relative clauses and multiple dependencies are the poster children for this type of model. ("My mother, who is a formidable cook, is thinking of opening a restaurant like the ones she and her family dined at in her youth in Europe.") For many sentences, memory-based models give similar predictions to surprisal-based models; but there are conflicts between the two:

> For memory-based theories, processing of X should be more difficult in the case with more dependents, due to the greater number of integrations, greater distance from X of early dependents and/or potential interference among dependents during retrieval. For expectation-based theories, on the other hand, the additional information obtained from more preceding dependents implies that the expectations of the comprehender regarding when X will be encountered and what input will instantiate it will generally be sharper and more accurate; thus there should on average be less processing difficulty at X than in the situation with fewer preceding dependents.[119]

As the field is still young, it is as yet an open question whether one account will eventually dethrone the other. According to computational psycholinguist Roger Levy (from whose work the above is drawn), it is likely in fact that both models could contribute to an understanding of how humans actually process language; he notes that one team currently working in the field has created a model that has "both prediction and verification components, which respectively yield surprisal-like and DLT-like processing difficulty gradients."[120] For our purposes, it is probably enough to remember that sentences with complicated dependencies are likely to contribute to the overall difficulty

---

[118] Roger Levy, "Memory and surprisal in human sentence comprehension" in *Sentence processing,* ed. Roger P. G. van Gumpel, New York: Psychology Press (2013): 82.
[119] Ibid: 100.

[120] Ibid: 108.

of language processing, and thus to the detriment of intelligibility when that language is sung.

However, it is worth noting that another approach to the problem of relative clauses has found that a surprisal-like effect applies to their processing as well. Psychologists Florencia Reali and Morten H. Christiansen compared the processing of object-relative sentences (“The reporter that the senator attacked admitted the error”) and subject-relative sentences (“The reporter that attacked the senator admitted the error”)—specifically those with pronouns in the second noun-phrase position (“The reporter that he attacked admitted the error”/ “The reporter that attacked him admitted the error”). Confirming earlier experiments, they found that object-relative sentences are much easier to process than subject-relative sentences. The twist is that the corpus they used had a vastly larger number of the former; object-relative sentences are much more common.[121] They conclude that, just as more common words are easier to process, so are more common sentence structures. They “look toward a model of sentence processing in which the system is influenced by statistical information defined at multiple levels of abstraction.”[122] In other words, in which frequency and predictability influence processing difficulty both at the level of the word and of the grammatical structure. It is probably enough for a composer to note that any sentence structure that seems odd or unusual is more likely to create problems with intelligibility—again, a new piece of information for the text-setting composer.

---

[121] Florencia Reali and Morten H. Christiansen, “Processing of relative clauses is made easier by frequency of occurrence,” *Journal of Memory and Language* 57 (2007): 17.
[122] Ibid: 19.

Discussion

How can all of this discussion of word frequency and length, of surprisal and entropy, of relative clauses and dependencies, be used in the toolkit we are assembling to analyze and maximize the intelligibility of classically sung English?

The overarching lesson is that words that are less predictable in their context are going to be more vulnerable to mishearing. Composers or analysts could, if so inclined, avail themselves of the available tools of computational psycholinguistics (such as the Roark parser) to calculate the relative surprisal and/or entropy of words in a text; or they could rely on an intuitive grasp of the concepts. (Either way, this is a new tool for musicians—these ideas have not hitherto been put in the service of composition!) Words or sentence structures that are found to be less predictable, if it's important that they be understood, can be set with special attention to range, repetition, rhythm, accompaniment density and the other aspects of text setting that have been found to affect intelligibility.

These five chapters have examined existing literature bearing on the intelligibility of classically sung text from the disciplines of music cognition, acoustics, linguistics, phonetics and psycholinguistics (including computational psycholinguistics). Each of these disciplines has its own priorities and perspectives on the issue of intelligibility; the task has been to tease out and weave together the threads that are relevant to our concerns as musicians. We have teased out an array of rubrics with which to analyze a text for its likely intelligibility, and to set that text so as to minimize the interference of the musical setting. In Chapter Six, I will assemble this interdisciplinary toolkit in one place, developing a set of guidelines for maximizing the intelligibility of sung English. In Chapters 7 - 13, I will analyze several selections from the worlds of opera, art song, operetta and musical theatre to see how well the guidelines match up with the practices of experienced composers.

Chapter 6: The Wolfson Heuristic for Analysis of Textsetting

If we combine all of the information we have gleaned from the previous chapters into one analytical tool, what would that look like? What would the procedure be for analyzing a text prior to setting it? What could the critical apparatus tell us about a completed text setting? In this chapter I outline what I am calling the Wolfson Heuristic for Analysis of Textsetting (WHAT).

The first step in using WHAT will be to analyze the text as text.

- Scan the text for difficult or unusual sentence structures, which can make the whole task of language processing more difficult, particularly if combined with syntactically complex music. Since the evidence for surprisal-type effects in the comprehension of sentence structure is much more sparse than for the equivalent effects in words, we will restrict ourselves to noting obviously odd or puzzling grammatical structures.
- Look for words that have high surprisal. These can be archaic or obscure words, or simply words that are unexpected in their contexts. Since the tools used by linguists are not designed for use by the layman, I will use a tool that is: the (free) list of the top 5000 most frequent words/lemmas[123] from the 450 million word Corpus of Contemporary American English, available

[123] A *lemma* is the dictionary, or canonical form of a word with multiple morphologies; “be” is the lemma that includes the words “is,” “are,” and “am.”

from http://www.wordfrequency.info/free.asp.[124] Words that are near the bottom of the list can be considered to have higher surprisal than words near the top; words that do not appear on the list at all can be considered to have higher surprisal still. Allowance must be made, however, for names and other words used consistently in a certain context, which will lower their surprisal.[125]

Another readily available online tool that can be useful in getting an idea of the relative frequencies of words is the Google Books Ngram Viewer (https://books.google.com/ngrams). This can be used to help recognize words that are unusual in their context; for instance, as a continuation of *walk the*, one is 1,000 more likely to encounter the word *streets* as the word *cat.*

- Look for words that increase or have high entropy. These are words that make it less clear what's coming next. Does the word narrow in the range of possible continuations, or expand it? Short of using the Roark Parser or a similar tool, this will be the most difficult to calculate accurately. Since entropy is largely equivalent to surprisal for most contexts, in the analyses following, I will omit this category unless there is an obvious effect.

---

[124] Lists of the 20,000 and 60,000 most common words are also available from http://www.wordfrequency.info/purchase.asp. I have made use of the list of 20,000 words in the following analyses, to distinguish the uncommon word from the truly obscure one.

[125] The familiarity (and thus surprisal) of certain words, phrases and even sentence structures may of course vary over time; words that may have been low-surprisal when set by Purcell (for instance) could require more cognitive effort from modern audience members. As the question of how intelligibility varies over time is beyond the scope of this paper, I have used contemporary word-frequency lists and the most recent data available from the Google Ngram Viewer.

• Look for words that are easily confused with other words by virtue of having large cohorts, especially if context does not help to narrow down the possibilities. If a word's cohort includes words much more common than itself, that will add to the potential difficulty. I have, somewhat arbitrarily, chosen a cohort size of twelve common words as the trigger for labeling a word's cohort as large. Many words, of course, have cohorts much, much larger, but it was surprising to me how many words have very few competitors.

I have been unable to find an easily usable tool that accurately lists words by their phonetic beginnings. Luckily, with a little bit of work it is not difficult to use a dictionary or other word list for this; most (though not all) words that begin with the same sounds will also begin with the same letters.[126] The most telling part of the analysis on this level is choosing what portion of the word to consider the cohort for. If the word is "hand," for example, (/hand/), are we considering words that start with /h/, words that start with /hæ/, or words that start with /hæn/? This will depend on the duration of the word in its musical setting; the longer a given vowel is held, the less important any following consonants will be to the cohort's

---

[126]After trying a paper dictionary, I eventually turned to an online tool designed for word games: Litscape.com's Word Finder tool at http://www.litscape.com/word_tools/starts_with.php. Litscape's corpus of about 138,000 words includes both common words and less common ones, but it also includes every possible form of the words on the list. To find the cohorts sizes of the words in the following analyses, I scanned the tool's results after entering letters for the first two phonemes of the word, and compiled a list of the basic forms of the common words included in those results. In calculating the cohort size for the word *death,* for instance, I included the word *deft* but not *deftly, deaf* but not *deafen.*

competition. In the analyses, I consider the cohort through the 1st vowel unless otherwise specified.

• Look for words that are potentially ambiguous by virtue of their phonetic properties.

—Since voiced stops (/b/, /d/ and /g/) are easily confused with one another, we want to be on the lookout for words in which substitutions of one voiced stop for another create new words, essentially increasing the size of the cohort. This is more likely to be problematic when the voiced stop is the first sound in the word. It is not difficult to examine these substitutions on the fly, as it were; the word "build" could be confused with "gild," but there is no word "dild." "Bug" could be confused with "dug," but there is no word "gug." However, "bud," "dud," "bub," and "dub" are also potential competitors.

—The same is true of unvoiced stops (/p/, /t/, /k/): we want to be on the alert for words in which substitutions of one unvoiced stop for another create new words, especially when the unvoiced stop is the first sound in the word.

—Words beginning with unvoiced fricatives (/f/, /θ/, /s/, /ʃ/, /h/) are potentially dangerous to intelligibility, especially at loud volumes or on high pitches.

—Words beginning with consonants and followed by front vowels (/i/, /e/, /ɛ/, /æ/)[127] can also be difficult to identify, probably because consonants are harder to identify correctly when followed by these vowels.

—In words that contain a back vowel (/u/, /ɔ, /ɑ/) followed by a consonant, that consonant may be harder to identify.

• Check to see whether any of the words singled out by the above steps are monosyllabic; these are the most likely of all to be confused with other words.

Any given word may be supported by its context, of course, possibly to the point that any phonological ambiguity or cohort size is moot. And any given word may fall into more than one of these categories and be considered to run an especially high risk of unintelligibility. In addition, of course, it will be a good idea to check for words that are simply difficult to sing by their nature (ala Rosenhaus and Cohen): words that feature densely packed consonants or frequent sibilants, for example, will pose a challenge to even the most adept singer.

Once we have identified the words that are at greatest risk for either being misheard or causing processing bottlenecks, the next step in using WHAT is to analyze the musical setting of the text for its treatment of these “difficult” words. (In the case of a composer newly setting a text, this will of course be an analysis of potential possible

[127] /a/ is also listed in some texts as a front vowel; however, it is generally assumed only to occur in English only as the first part of the diphthong /ai/ (as in *kite*). It is unclear whether, when the first vowel of this diphthong is extended in a sung setting, /a/ is the vowel that is actually sung, or whether a back vowel such as /ɑ/ is substituted; therefore, I will be omitting /a/ from this category of the analysis unless the setting is sufficiently speechlike in tempo and range that it seems appropriate.

settings.) Note that the analysis outlined below does not need to be "pro-intelligibility" on every count for a text to be intelligible; the premise of this thesis is that problems with intelligibility can arise when "difficult" words are married to musical settings that exacerbate them.

- Does the metric setting conflict with the text's word and sentence stresses?
- Are word-groups (as defined by Thomson) set so their boundaries are obscured?
- Does the tessitura sit above the treble staff, where most vowels are unrecognizable, and consonants (except for the nasals and liquids) also decline in intelligibility?
- How are "difficult" words set (high surprisal/entropy, ambiguous, potentially phonetically ambiguous or large cohort), especially words that fall into more than one of these categories?
- Are high surprisal words given extra time compared to surrounding words, as they would be in speech?
- Does any range-related perception of vowel substitution that takes place make a word sound like a different word, or a nonword? Besides the effects of the fundamental overtaking the vowel's formant above the treble staff, remember that "as the fundamental rises, "i" will begin to sound like "I" and then "Ɛ;" "I" will begin to sound like "Ɛ" and then "æ;" and "Ɛ" will begin to sound like "æ." " Thus the vowels /i/ and /u/ will be most identifiable between the notes B3 and F#4, the vowel /I/ between F#4 and C#5, /ɛ/ between C#5 and F#5, and /ɑ/ between F#5 and B5. /i/ begins to

be interpreted as /I/ at (D5) and as /Ɛ/ at (F#5); /I/ begins to sound like /Ɛ/ at D5 as well; /Ɛ/ begins to be interpreted as /æ/ as low as (F#4). All of this will apply most directly to words or syllables that are held for a relatively long time, increasing the isolation of the vowel and lessening the impact of formant transitions on its perception. (See Figure 1.)

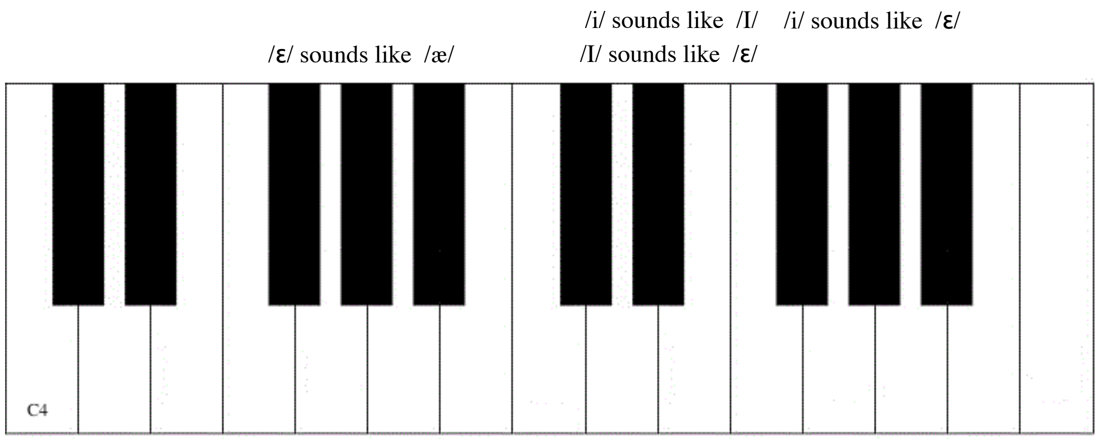


Figure 2, Range-related Vowel Substitution

- Are words possibly ambiguous because of unvoiced fricatives [s], [ʃ], [f], [θ] set on high or loud notes?
- Does chunk density exceed 3-5 musical chunks per 2 seconds accompanying "difficult" words?
- Does the tempo or melismatic setting make it difficult to retain the beginnings of words in working memory long enough to reach their recognition point? Are melismatic settings, especially of monosyllabic words, followed by syllabic ones?

- Is harmonic priming working for or against intelligibility? Instances in which a word is set to an unexpected harmonic event, such as an evaded cadence or a deceptive cadence, may contribute to the difficulty of processing the word, especially when it is already “difficult” for another reason.
- The repetition of “difficult” words may assist in making them intelligible. Does the setting take advantage of this effect?

In the following seven chapters, I use WHAT to analyze settings of English text from opera, operetta, art song and musical theatre. Some of the pieces I chose have a reputation for being difficult to understand, and were recommended to my attention in an informal survey of singer friends; others are songs I have long been familiar with, and was curious about. Five of the seven were written for soprano, as this is the voice type for which setting text intelligibly seems to be most difficult, but I have included a selection for tenor and one for baritone to give a more varied picture. The analyses are presented in the order in which I completed them, rather than in any systematic way; each of them has a slightly different emphasis, as each has different features, posed different analytical challenges, and yielded different insights into the intelligibility of the setting.

Chapter 7: Analysis, "Dido's Lament."

"Thy hand, Belinda/When I Am Laid In Earth" (Dido's Lament) from *Dido and Aeneas* by Henry Purcell and Nahum Tate.

This recitative and aria, the latter famously composed over a ground bass consisting of a descending chromatic tetrachord, is from Purcell's only opera. The exact date of composition is unknown, but it seems to have been written between 1685 and 1689. The only known performance of the opera during Purcell's lifetime was at a girls' boarding school; it is not known whether that was the original version, and the surviving scores calling for male voices were later revisions, or whether it was originally composed for a court performance (like its model, John Blow's *Venus and Adonis*) and subsequently transplanted to the school.[128]

The aria was recommended by two different people during an extremely informal poll of singer friends, in which I had asked for suggestions of songs known to be difficult to make intelligible. This piqued my interest, as it was a piece I knew that had never struck me as particularly problematic in that regard. Surely it would not be so famous if it were difficult to understand?

The text of the recitative and aria (for Dido, a soprano):

> Thy hand, Belinda[129];
> Darkness shades me:
> On thy bosom let me rest:

---

[128] *The New Grove Dictionary of Opera,* "Dido and Aeneas, " Oxford University Press, http://www.oxfordmusiconline.com.proxy.libraries.rutgers.edu/subscriber/article/grove/music/O006883, (accessed February 20, 2017).

[129] Some versions of the score have this as "Thy hand, my Anna." The name Anna for the character of Dido's handmaiden, to whom the recitative is addressed, is from a posthumous revision of the score in which the name of Dido's sister from *The Aenead* was substituted for Tate's choice.

> More I would, but Death invades me.
> Death is now a welcome guest.
>
> When I am laid,
> Am laid in earth,
> May my wrongs create
> No trouble, no trouble in thy breast;
> Remember me, remember me,
> But ah! Forget my fate.
> Remember me, but ah!
> Forget my fate.

Analysis of this text using the methods outlined in Chapter 6 yields the following.[130]

Sentence Structures

On thy bosom let me rest is an inversion of normal sentence order (as opposed to "let me rest on thy bosom"). *More I would* (meaning, presumably, "I would say more") is a definitely unusual structure, possibly enough to interfere with comprehension even if intelligibility is not otherwise compromised. The repetition of *am laid* obscures the structure of the sentence *When I am laid in earth, may my wrongs create no trouble in thy breast.* (It is more problematic than the repetition of *no trouble* because of the breaking of the word group "I am.")

Surprisal

The first word of the recitative, *thy,* is an uncommon word, not appearing at all in the list of the 5,000 most frequently used English words. Moreover, this is the first time it

[130] When heard in the context of the opera, audiences will have as context the fact that Dido's lover Aeneas is leaving her. Presumably this will give a slight boost in their cohorts to words and phrases that reflect anguish.

is used in the libretto of the opera. However, it has a very small or even nonexistent cohort: once we have heard the vowel, there is no other word it can be (except its cognate "thine"). *Bosom* is also an uncommon word (again, its frequency is less than any of the top 5,000 words), and there are other, more common possibilities following the words *on thy*; "honor," for instance, is at number 2,096 on the list. *Invades*, while more common (4,562 on the list), is definitely an unexpected word in the context.[131]

Cohort size

The following words have cohort sizes (after the first vowel, except as noted below) of twelve common words or more: *hand*, *more*, *but*, *Death*, *Earth*, and *remember*. Please see the Appendix for discussions of the cohorts and recognition points of each word.

Phonetic ambiguity

•Words starting with /b/, /d/ or /g/ in which substitutions among those three phonemes create other words include *guest, breast*, and *but*. Of these, *guest* could be confused with *best*, *breast* with *dressed*, and *but* with *gut*. Context makes all of these unlikely.

•The only word containing /b/, /d/ or /g/ in a non-initial position that might be vulnerable to substitution is *hand*; it could be mistaken for *hang*. (This would be very

---

[131] The Google Ngram Viewer tells us that the word "death" is most often followed by conjunctions and prepositions (with "of" being the most common next word, at approximately .002% of the corpus). "Death" is followed by a verb, any verb, slightly less often than by the word "of," and followed by "invades" constitutes only about .0000007% of the corpus.)

unlikely in speech, but given the lack of nasalization on vowels before nasal consonants in singing, it is a possibility here.)

•The only possible word ambiguity resulting from substitution among /p/, /t/ and /k/ is hearing *fate* as *fake.*

•Words that begin with unvoiced fricatives are vulnerable to substitution at high pitches and/or high volumes. The only one such here is *hand,* whose cohort becomes even bigger if the first phoneme is ambiguous; however, it is not set at a high pitch or volume.

•Words beginning with consonants followed by front vowels include *hand, shades, me, let, me, rest, death, welcome, guest, when, laid, may, create, breast, remember,* and *fate.* Since this is sixteen out of a total of thirty-seven words in the text, we will not use this to single out individual words, noting only that the preponderance of this type of word might decrease the overall level of intelligibility of the piece.

•Words containing a back vowel followed by a consonant include *darkness, more, wrongs,* and *forget.* These words may be more at risk of mishearing.

"Difficult" words

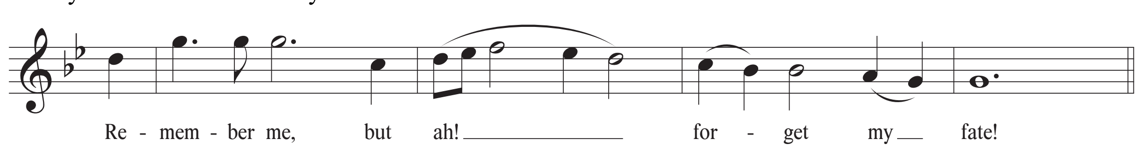
Hand appears as potentially difficult three times in this analysis, *more* and *but* two each. *Death, earth, remember, guest, breast, fate, darkness, wrongs, forget* all appear once. The monosyllables among this group are *hand, more, but, Death, earth, guest, breast, fate,* and *wrongs.* We will pay special attention to these words in the analysis of the musical setting below.

Analysis of the musical setting is as follows. Since there is no authoritative version of this score, I have referenced both the 1841 Academy version and the 1880 Purcell Society version where appropriate.

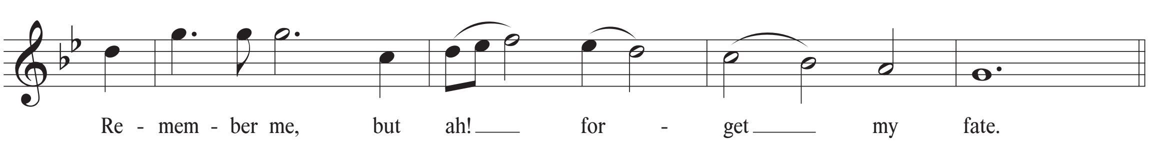
Metric setting

The word *forget* is set with the first (normally unstressed) syllable on the downbeat in the Purcell version of the score, and midway through the second beat in the Academy version. (See Examples 7.1a and 7.1b, respectively.) The Purcell version's setting is a metric mismatch, with a greater risk of unintelligibility. It is difficult to know whether the metric setting of *"More I would"* is normative, since it's such an odd sentence structure; but *more* looks to be the word that should be stressed, and it's not. Again, this could potentially interfere with the decoding of an already difficult-to-decipher sentence. Other than that there are no obvious mismatches of metric and musical stress. (See Example 7.2.)

Example 7.1a. "Dido's Lament" from *Dido and Aeneas* by Henry Purcell and Nahum Tate, m32-35, Purcell Society score. Vocal line only.



Example 7.1b. "Dido's Lament" from *Dido and Aeneas* by Henry Purcell and Nahum Tate, m32-35, Academy score. Vocal line only.

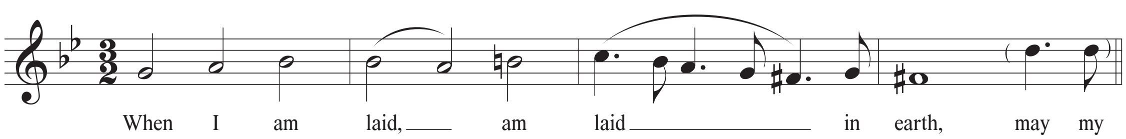


Example 7.2. "Dido's Lament" from *Dido and Aeneas* by Henry Purcell and Nahum Tate, m6-8 (all versions). Vocal line only.

Word Groups

As noted above, the word group “I am” is broken by the repetition of “am laid,” at a potential cost to intelligibility. (See Example 7.3.)

Example 7.3. “Dido’s Lament” from *Dido and Aeneas* by Henry Purcell and Nahum Tate, m15-18 (all versions). Vocal line only.



Tessitura

With the exception of the final “remember me,” of which the last three syllables are set on G5, the entire setting lies between C#4 and F5. The first syllable of this final “remember me” is on D5; thus, despite the higher tessitura of the other three syllables, we will probably recognize the repetition, and it will certainly be confirmed at the end of the measure when we hear the word “but” on C5, signaling that this is a textual repetition. (See Example 7.1.)

Settings of difficult words

The potentially difficult words previously identified are: *hand, more, but, Death, earth, remember, guest, breast, fate, darkness, wrongs, forget.*

There is nothing noteworthy about the settings of *hand, but, Death, earth, guest, breast,* or *fate. More* is notable for its metric setting, already discussed (see Example 7.2); *forget* is set with a metric mismatch, already discussed.

Wrongs is set to a three-note melisma, taking up two beats. While this is certainly diminishing its intelligibility compared to a syllabic setting, this is probably a minimal effect.

Darkness, in the Purcell Society score, has a five-note melisma on the first syllable, taking up all but one sixteenth note of the 4/4 bar of recitative, with the second syllable taking up that last 16th note. (In the Academy version, the first syllable is a three-note melisma over three quarter notes.) Collister and Huron's studies of melismas did not include this case, in which the first syllable of a bisyllabic word is set melismatically but the second is set syllabically; however, the Cohort Model suggests that the syllable "ness" will force the selection of *darkness* from among its competitors even after a melisma, at no cost to intelligibility. (See Example 7.4.)

Example 7.4. "Dido's Lament" from *Dido and Aeneas* by Henry Purcell and Nahum Tate, m1-5, Purcell Society score. Vocal line only.

High surprisal words

Thy is set on short notes (eighth notes), thus potentially adding to the difficulty of decoding it. *Bosom* is set as the longest word in its sentence, more accurately reflecting the intelligibility-boosting longer durations of high-surprisal words in speech. (See Example 7.4.) *Invades* is set as two eighth notes; the intelligibility of this word maybe downgraded to some extent, although since this is in a recitative, the singer is free to lengthen these durations (and in many recordings of the piece this does in fact occur). (See Example 7.2.)

Perception of vowel substitution

When, breast, and *forget* all feature an /ɛ/ vowel above F#4, although *when* and *breast* are only slightly above. The potential mishearing of /ɛ/ as /æ/ would give us the nonwords *whan, brast,* and *forgat.* The first two statements of *Remember me* have both /i/ and /ɛ/ on D5; the potential mishearing there would be something like *rimamber mih*. The third statement of *remember me*, with all but the 1st syllable on G5, would probably be perceived as something like *rehmahmber meh*. (See Example 7.1.) However, the only one of these in which the vowel is held in isolation for a relatively long time is *breast*; this is probably the only word in the aria vulnerable to mishearing specifically because of this effect.

Unvoiced fricatives

We are looking here for unvoiced fricatives on loud or high notes. No dynamics are specified in the score, but the only high note is a setting of *remember me*, which does not contain any unvoiced fricatives; the only possibly qualifying word here is the third appearance of *forget* (on E♭5). (See Example 7.1a.) This could downgrade intelligibility, but the obvious textual repetition should make this effect minimal.

Chunk Density

The accompaniment moves in half notes at a stately tempo; there is no risk of overloading the auditory input.

Recognition Points

If we assume that the tempo marking Larghetto implies a metronome marking of 60 bpm, then no single word lasts more than the two second limit of working memory. The only melisma of more than two notes is that on the 2nd appearance of the word *laid.* We recall from Collister and Huron 2008 that a melismatic setting after a syllabic one does not increase recognition; because the repetition of *am laid* is breaking up the word group *I am,* there is likely to be some confusion at this spot anyway, and the five note melisma on *laid* may exacerbate the problem. (See Example 7.3.)

Harmonic priming

Of the "difficult" words already identified, *hand, but* and *Death* (in its first appearance) occur during a sustained chord, and will not be affected by harmonic priming. The same is true of *Darkness,* in the second measure, which appears on a V/iv chord, not too unusual following the first measure's minor tonic triad. *Death*'s second appearance changes the chord from C major to c minor (the latter of which acts as a pivot from the recitative's opening c minor to the g minor of the aria); the modal mixture certainly is a harmonically surprising event in the context, and might impede recognition to a degree, although this will likely be offset by its proximate repetition. *Guest* is set at an entirely expected half cadence, as are *earth* and *breast*; *remember* is set on the resolution of an authentic cadence. None of these should be impeded by harmonic priming effects.

The first appearance of *fate* occurs accompanied by a V^6 chord, in the second measure of the passacaglia's chromatically descending bass line. The effect here is somewhat half-cadential, which is interestingly unexpected shifted as it is away from the actual location of the authentic cadence at the end of the bass line. (See Example 7.5.)

Example 7.5. "Dido's Lament" from *Dido and Aeneas* by Henry Purcell and Nahum Tate, m28-35, Academy score. Vocal line and figured bass.

However, the chord itself is not unusual for its location in the bass line, and as we've heard the pattern four times already it will likely not be sufficiently unexpected to contribute significantly to any decrease in intelligibility. In any case, we hear the word again four bars later as part of a perfect authentic cadence, which should reinforce it. Harmonic priming is unlikely to be an impediment to intelligibility here.

Conclusions

There are two spots in this aria in which difficulties converge, raising the possibility that intelligibility will suffer. One is the phrase *More I would,* in which both the unusual sentence structure and the metric setting might combine to undermine intelligibility; the other is *When I am laid, am laid in earth,* in which the broken word group and the melisma on the 2nd *laid*, as well as its delayed recognition point, might also place an undue burden on the listener.

There are numerous small potential strains on intelligibility: the high surprisal word *thy* is not given extra time relative to the surrounding words; one appearance of the word *forget* (beginning with an unvoiced fricative and containing a back vowel followed by a consonant, both intelligibility-risking) is set in one version on a moderately high, loud note, with a metric mismatch to boot; the word *breast* is sustained long enough that the perceived vowel substitution could affect its intelligibility. A large percentage of the text also consists of consonants followed by front vowels; since these are potentially more difficult to decode, this preponderance could tax the overall decoding process.

The difficulties of making this aria intelligible certainly add to the not inconsiderable vocal difficulties of performing it. This analysis justifies the rhythmic choices that performers of this song often take in the recitative, giving high surprisal words more time; it also potentially sheds light for an interested singer on the choice of which version of the aria to perform.

Chapter 8: Analysis, "Green Finch and Linnet Bird."

"Green Finch and Linnet Bird" (first two stanzas) from *Sweeney Todd* by Stephen Sondheim and Hugh Wheeler.

Sweeney Todd, originally produced on Broadway in 1979, is one of Sondheim's most "classical" sounding scores. It includes an opera pastiche, a recurrent quote from the Dies Irae, and a sophisticated harmonic language, and has been performed by many opera houses in addition to its musical theatre productions.[132] "Green Finch and Linnet Bird" is sung by Johanna, the ingenue soprano; she has been imprisoned in her room by her guardian, and is longing for freedom.

I have wondered about how this song could possibly be intelligible to an audience ever since I music directed a production of *Sweeny Todd* in summer stock over twenty years ago. It is chock full of unfamiliar words, and sung by a soprano. That ought to be a recipe for disaster. Yet somehow it is not. Why?

The lyrics of the first two stanzas of the song:

Green finch and linnet bird, nightingale, blackbird,
How is it you sing?
How can you jubilate, sitting in cages
Never taking wing?
Outside, the sky waits, beckoning, beckoning
Just beyond the bars.
How can you remain, staring at the rain,
Maddened by the stars?[133]

---

[132] http://www.sondheimguide.com/sweeney.html#HGO, (accessed 4/18/17).
[133] The song continues:
How is it you sing anything?
How is it you sing?
Green finch and linnet bird, nightingale, blackbird,
How is it you sing?
Whence comes this melody constantly flowing?
Is it rejoicing or merely halloing?
Are you discussing or fussing

Analysis of this text using the methods outlined in Chapter 6 yields the following.

Sentence Structures

The opening sentence of the song, which begins with the species names of four birds used as a direct address, is unusual, although its comprehension will presumably be facilitated somewhat by the intended staging, in which Johanna looks out her window at a bird seller's birds, and the dramatic analogy drawn between her imprisonment and theirs. The rest of the song consists of simple questions and imperative sentences.

Surprisal

Words that do not occur in the 5,000 most frequent words in the English language include:

finch, linnet, nightingale, blackbird (although *black* and *bird* are both on the list), *jubilate, beckoning,* and *maddened.* Of these, according to the Google Ngram viewer, *finch, nightingale, blackbird, beckoning,* and *maddened* are a couple of orders of

---

Or simply dreaming?
Are you crowing? Are you screaming?

Ringdove and robinet, is it for wages,
Singing to be sold?
Have you decided it's safer in cages,
Singing when you're told?
My cage has many rooms, damask and dark
Nothing there sings, not even my lark.
Larks never will, you know, when they're captive.
Teach me to be more adaptive. Ah,
Green finch and linnet bird, nightingale, blackbird,
Teach me how to sing.
If I cannot fly,
Let me sing.

magnitude less frequent than, say, *bird*; *linnet* and *jubilate* are one or two orders of magnitude even less frequent and can be considered truly obscure. (In addition to those two words, *finch, nightingale,* and *blackbird* do not turn up even in the list of 20,000 most frequent words.) As indicated previously, many of the obscure bird names will be given context by the staging, and thus be more identifiable. Not only that, but the fact that the word *bird* turns up twice in the first line (if we count the compound form blackbird) will probably assist with comprehension. *Jubilate, maddened,* and *beckoning* are cognates of slightly more usual forms (*jubilation, madden* (11,587), *beckon* (8,964)).

Besides frequency effects, we do not often say that the sky *waits*.[134] This bit of poetic imagery will likely come at some possible cost to intelligibility.

<u>Cohort size</u>

The following words have cohort sizes (after the first vowel) of twelve common words or more: *finch*, and, *bird*, *can*, *sitting*, *wing*, *outside*, *beckoning*, *just*, *beyond*, *remain*, *staring*, *at*, *rain*, and *by*. Note that this is sixteen out of the forty-four words in the excerpt. Please see the Appendix for discussions of the cohorts and recognition points of each word.

[134] According to the Google Ngram Viewer, *waits* follows *sky* about 1/1000th as often as *at*, the 10th-place finisher.

Phonetic ambiguity

•Words starting with /b/, /d/ or /g/ in which substitutions among those three phonemes create other words include only *bird* and *bars*, which could be heard respectively as *gird* and *gars.* Both of these are less common.

•There are no words in this excerpt containing /b/, /d/ or /g/ in a non-initial position that could cause confusion with another word because of substitution among these consonants.

•There is some possible ambiguity resulting from substitution among /p/, /t/ and /k/: *sitting* could be heard as *sipping*; *taking* could be heard as *caking*; *sky* could be heard as *sty*; *staring* could be heard as *scaring* or *sparing*; *stars* could be heard as *scars* or *spars*.[135]

•Words that begin with unvoiced fricatives include *finch, how, sing, sitting, sky, staring, stars.* However, with the possible exception of *stars,* none of these is set particularly high or loudly.

•Words beginning with consonants followed by front vowels include *green, blackbird, can, cages, never, taking, waits, beckoning, beyond, remain, staring, rain,* and *maddened.* Again, this is nearly half the words in the piece; we will consider this effect only as it adds to other contributions to unintelligibility.

•Words containing a back vowel (/u/, /ɔ, /ɑ/) followed by a consonant include *beyond, bars,* and *stars.* These words may be more at risk of mishearing.

---

[135] The effect of alliteration on consonant perception has not been studied, but it seems possible that either correct hearing or mishearing of the first word of an alliterative pair such as *staring* and *stars* would influence the perception of the second of the pair.

"Difficult" words

Staring appears four times as potentially difficult twice in this analysis. Words that appear three times: *beckoning, stars.* Words that appear twice: *finch, nightingale, blackbird, can, beyond*, *bars, sitting, taking*, *rain, by, beyond, maddened* and *remain.* Words that appear once: *bird, and, wing*, *outside*, *just*, *remain*, *staring*, *at*, *sky*, *cages, never, taking, waits, green, linnet* and *jubilate.* This is the majority of the words in the excerpt. The monosyllables on this list are: *stars, finch, can, bars, rain, by, bird, and, wing, just, at, sky, waits, green.* Note that this is a majority of the words in the excerpt.

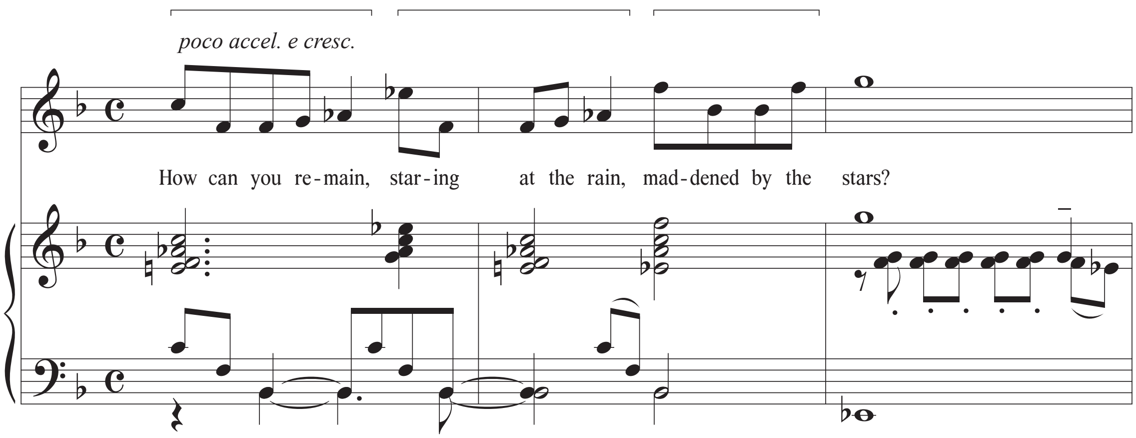
Analysis of the musical setting is as follows.

Metric setting

The words *green finch,* as a noun phrase, would be a spondee; Sondheim has set them as though they were a trochee, with the word *finch* relatively unstressed and shorter than *green.* While this is an odd setting, and makes the uncommon word *finch* potentially less intelligible, there actually is a type of bird called a "greenfinch." This word is, in fact, a trochee. I speculate that the published score has a typo, in essence, and is missing the hyphen between the two syllables that would indicate that *green* and *finch* are two syllables of the word *greenfinch.* (See Example 8.3, below.) Still, there is a potential for intelligibility decrease here.

The *phrase staring at the rain* is set with its stressed syllables on weak beats, and the unstressed word *at* on a downbeat. However, the three-beat phrase mimics the previous one, which is set with its stresses in the expected places; the first syllable of *staring* also features a significant melodic accent, as does the first syllable of the next phrase, *maddened by the stars.* The effect is one of a 3+3+2 hypermeter across the two bars of common time, which is reinforced by rhythm of the accompaniment; looked at this way, the metric mismatch disappears. (See Example 8.1.)

Example 8.1. Measures 13-15, "Green Finch and Linnet Bird" from *Sweeney Todd* by Stephen Sondheim.



Word Groups

The integrity of word groups is maintained throughout the setting.

Tessitura

The excerpt sits within the treble staff except for the word *beckoning,* whose 2nd syllable is set on a G♭5 both times (see Example 8.2), and the word *stars*, which is set on G5 (see Example 8.1). Both of these words have been singled out as potentially difficult, and these settings may exacerbate the issue. However, *beckoning* is repeated immediately, which has been shown to increase intelligibility; and *stars* should get a boost in intelligibility from its rhyme with *bars* (while *bars* is itself a potentially difficult word, it is set on C5 and is certainly less vulnerable than *stars*).

<u>Settings of difficult words</u>

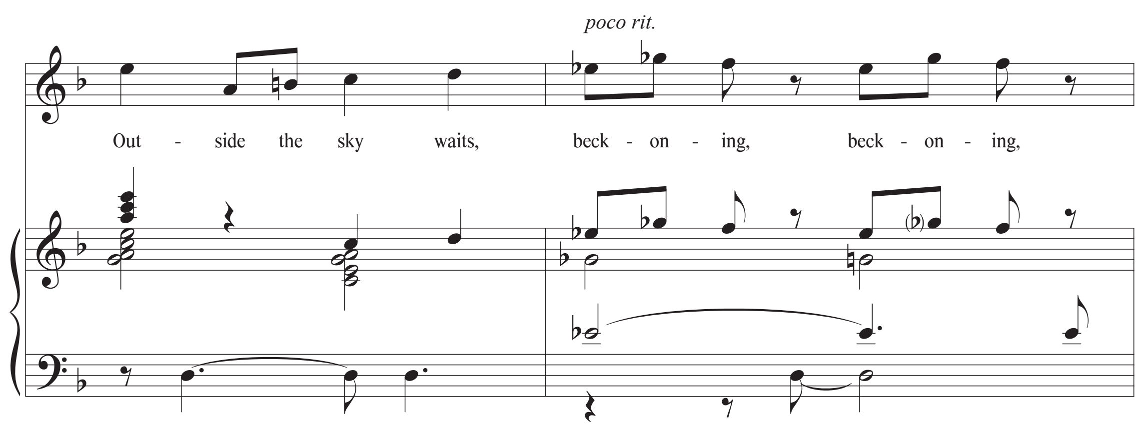
Since nearly every word in the excerpt appears as potentially difficult in this analysis, I will not attempt to single out any particular word's setting.

<u>High Surprisal Words</u>

Finch (or *greenfinch*), *linnet, nightingale, blackbird, jubilate, beckoning, maddened* and *waits* were all tagged as potentially high suprisal words. *Nightingale* and *blackbird* are given settings of two beats each, which at the given tempo feels adequate; so is *jubilate,* which is however a much more obscure word. *Linnet,* equally as obscure, only gets one beat and may suffer (although it's arguable that *linnet bird* might be heard as a compound, *linnet-bird,* since *linnet* is so obscure anyway).

Beckoning is set in a measure marked *poco rit.* While this is probably intended to contribute to the sensuous chromaticism of the setting (a reference to tropes of Romantic exoticism, a form of cultural word-painting), it also has the effect of giving a high-surprisal word a little extra time, increasing its intelligibility. (*Waits,* which falls on the last beat before the *poco rit.*, might in practice also gain a little extra time and intelligibility thereby.) (See Example 8.2.) Conversely, the word *maddened* is part of a passage marked *poco acc.*; especially given that, with *by,* it creates a cluster of consonants that will be difficult to enunciate clearly, this word's intelligibility is likely to be downgraded by its setting. (See Example 8.1, above.)

Example 8.2. Measures 9-10, “Green Finch and Linnet Bird” from *Sweeney Todd* by Stephen Sondheim.



Perception of vowel substitution

The only words vulnerable to vowel substitution held long enough for the effect to be potentially noticeable are *sing* and *wing,* which could be perceived as the nonwords *seng* and *weng.* Interestingly, however, *sing* is one of the few words not marked as potentially difficult by this analysis, and is supported by the context to boot; and *wing* is likely to be assisted by its rhyme with *sing.* Sondheim seems to have taken steps to minimize this effect.

Unvoiced fricatives

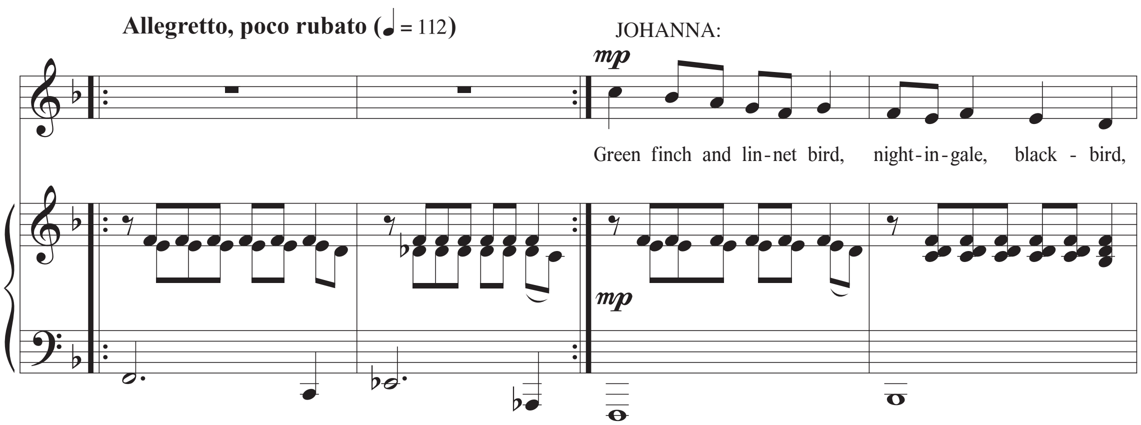
There is a slight potential difficulty with *how* in measure 5, which is set on F5 (although only marked ***mp***). Note that it is the second appearance of the word, however, with the first, one bar earlier, set safely on middle C; some boost to intelligibility will likely come from the repetition in parallel sentence structures. (See Example 8.4, below.)

Stars, marked ***mf*** on a G5, is perhaps a greater challenge to intelligibility, although the possible mishearings due to ambiguity of the unvoiced fricative are all with nonwords.

Chunk Density

At ♩ = 112, the eighth notes of the accompaniment would come close to overloading working memory if they were not chunked. However, they are: the vamp preceding the entrance of the vocal, which introduces the pattern that then becomes the accompaniment, effectively consolidates it as a chunk. (This is a common strategy in musical theatre.) There should be no risk to intelligibility from this factor. (See Example 8.3.)

Example 8.3. Measures D, E, 1 and 2, "Green Finch and Linnet Bird" from *Sweeney Todd* by Stephen Sondheim.



Recognition Points

The only words held longer than the two second limit of working memory (at a potential detriment to their processing) are the ends of three phrases: *sing, wing,* and *stars.* (They come in at 2.14 seconds.) Interestingly, *bars,* the other phrase-ending word, is notated as three beats long, while the others are all four beats; this will have the effect, if observed by the performer, of making *bars* slightly less difficult to process. Perhaps this in turn will give a boost to *stars,* which as the only note above the staff is uniquely vulnerable here to mishearing, by cementing in the effect of the rhyme.

Harmonic priming

Of the "difficult" words in this excerpt, the following coincide with a change in harmony: *how* (in measures 5 and 13); *sitting*; *never*; *outside*; and *stars.*[136] Of these, only *how* comes on a harmony that is at all unexpected: while the vi chord in measure 5 is not unusual, it is not quite the expected resolution of the V4/2 of IV chord that precedes it. Note that, as a word beginning with an unvoiced fricative on a high note, we have already seen that this word is vulnerable. (See Example 8.4) In measure 13, *how* coincides with the change, facilitated by a melodic common tone, from four bars of a V/ii dominant lock to a V/VII; this could also come at a cost to intelligibility, although as it is the third instance of the word in a parallel sentence structure, this seems unlikely.

---

[136] I am choosing to view the chords in measures 2 and 10 as prolonging the prevailing harmonies, and thus not including *nightingale* or *beckoning* in this list.

Example 8.4. Measures 3-5, "Green Finch and Linnet Bird" from *Sweeney Todd* by Stephen Sondheim.

It is also worth noting that *stars,* perhaps the most problematic word in the excerpt in terms of melodic setting, is aided by harmonic priming: it coincides with the entirely expected resolution of that V/VII to VII. (See Example 8.1.)

Conclusions

Sondheim uses heightened, slightly archaic language throughout much of *Sweeney Todd* (in combination with a musical language that shares a melodic and harmonic vocabulary with many operas). This could present problems for intelligibility, but at least in this excerpt the preponderance of unusual words is balanced by a context that will assist in their processing, and possibly their recognition. All of the unusual bird names are given a frame by the intended staging and use of the word *bird* as a combination form in the first lines; even the short settings of the high surprisal words *finch* and *linnet* may survive this.

There are still some intelligibility pitfalls in this setting. Like *Dido's Lament,* there is a large number of words in which consonants are followed by front vowels. A large proportion of the words in the text are flagged by the analysis as difficult in some way. The word *maddened* seems very likely to be compromised by its alread short setting being in the midst of an accelerando.

However, some of the potentially most difficult settings have been finessed to be less problematic than they might have been. *Beckoning*'s high tessitura is offset by its being immediately repeated and slowed down. The high surprisal word *waits* gains time from a *poco rit.* The potentially phonetically ambiguous *wing* is the second word of a rhyming pair whose first member is supported by context. *Stars,* probably the most challenging word in the piece, is supported by being the second word of a rhyming pair whose first member's duration is short enough to not challenge working memory, and possibly also by its alliteration with the word *staring*. Perhaps this goes some way toward explaining the success of this unusual Broadway aria.

Chapter 9: Analysis, "Sea-Snatch."

"Sea-Snatch" from *Hermit Songs* by Samuel Barber (on anonymous Irish texts of the 8^{th} - 13^{th} centuries).

Hermit Songs was commissioned for the soprano Leontine Price, and was first performed by her with the composer at the piano in 1953. "Sea-Snatch" is one of ten songs, all settings of anonymous texts by medieval Irish monks and scholars, "often on the margins of manuscripts they were copying or illuminating."[137] It has since become part of the standard English-language art song repertory, and is available in versions for both high and low voices. Several of the pieces, including "Promiscuity" and "The Monk and his Cat" have become stand-alone favorites.

"Sea-Snatch" is another gleaning from my informal poll of singer friends; it was not a song I was familiar with (unlike the two mentioned above). However, upon listening to some recordings of the piece, including Leontyne Price's, I decided to include it here. It is a musically dramatic and arresting setting—of which perhaps two words in five are intelligible. The analysis if of the original (higher) key.

The text of the song:

It has broken us, it has crushed us, it has drowned us,
O King of the starbright Kingdom of Heaven;
The wind has consumed us, swallowed us,
As timber is devoured by crimson fire from Heaven.
It has broken us, it has crushed us, it has drowned us,
O King of the starbright Kingdom of Heaven!

Analysis of this text using the methods outlined in Chapter 6 yields the following.

---

[137] Samuel Barber, *Hermit Songs,* (New York: G. Schirmer, Inc., 1954): front matter.

Sentence Structure

The first line, despite the commas, should be easily heard as three short sentences; however, the second line, a long direct address, is quite likely to be confusing, given that the repetition of the first structure will have primed us to hear another simple sentence. *The wind has consumed us* resumes the simple subject/verb/object sentence structure, but then *swallowed us* breaks the pattern, forcing us to hear the comma (as it were) and interpret *swallowed us* as a duplicate predicate to the subject *The wind (has)*, which is an unusual structure, possibly more difficult to process.

The fourth line (*As timber…*) constitutes a subordinate clause, a not uncommon structure. Even stacked prepositions, here *as, by*, and *from,* are not particularly unusual; however, with each one the neural processing cost grows, with possible consequences for intelligibility. Again, the switch to a different structure from the repetition of the simple sentences may come into play, despite the fact that the pattern has already been disrupted by *swallowed us.* Sentence structure may play a role in contributing to unintelligibility in these spots.

Surprisal

The only words that do not appear in the list of 5,000 most commonly used words in the English language are: *starbright* (although both *star* and *bright* appear), *devoured* (8,643), and *crimson* (12,498). *Starbright* does not appear in the 20,000 most commonly used words, either. These may contribute to unintelligibility.

The familiarity of most of the words notwithstanding, there is plenty of surprisal in this short text. After the words *it has*, the word *been* is the most common, at .002% of the trigrams in the Google Ngram Viewer's English corpus. By contrast, *broken* comes in at .0000026%, *crushed* at .0000001%, and *drowned* at .00000004%—these words are less common continuations of *it has* by a factor of between 1,000 and 10,000. Moreover, the 4-gram *it has broken us* brings up a null result in the Ngram Viewer, as do *it has crushed us* and *it has drowned us.* These phrases do not appear even once in the entire Google Books English corpus. The same is true of the 3-grams *wind has consumed, timber is devoured*, *devoured by crimson, by crimson fire* and *crimson fire from.* High surprisal indeed! The demands on mental processing here will be notable, almost certainly affecting intelligibility.

<u>Cohort Size</u>

The words that have cohort sizes of twelve or more common words (after the first vowel) are *has O, King, of, Heaven, wind, consumed, timber, devoured,,* and *by.* Please see the Appendix for discussions of the cohorts and recognition points of each word.

<u>Phonetic ambiguity</u>

•The only word starting with /b/, /d/ or /g/ in which substitutions among those three phonemes create another word is *drowned*, which could be heard as *browned.*

•The only word in this song containing /b/, /d/ or /g/ in a non-initial position that could cause confusion with another word because of substitution among these consonants is *starbright*, which could be heard as *starred right*.

•There is some possible ambiguity resulting from substitution among /p/, /t/ and /k/: *King* could be heard as *ping* (although presumably the word *Kingdom* would clear that up once it appeared). If *starbright* is heard as a compound of *star* and *bright* (not unlikely given its unfamiliarity), *star* is vulnerable to being heard as *scar* or *spar.*

Combinations of mishearings of the above two categories could also lead to *starbright* being heard as *scarred right* or *sparred right.*

•Words that begin with unvoiced fricatives include *has, starbright, Heaven, swallowed, fire* and *from.* (Of these, *swallowed*, on a G5 at ***f***, is most vulnerable to mishearing.)

•Words beginning with consonants followed by front vowels include *has, King, Kingdom, wind, timber, devoured, crimson, Heaven.* These words may be more at risk of mishearing.

•Words containing a back vowel followed by a consonant include *broken, us, crushed, drowned, of, consumed, swallowed, devoured,* and *from.* These words may be more at risk of mishearing.

"Difficult" words

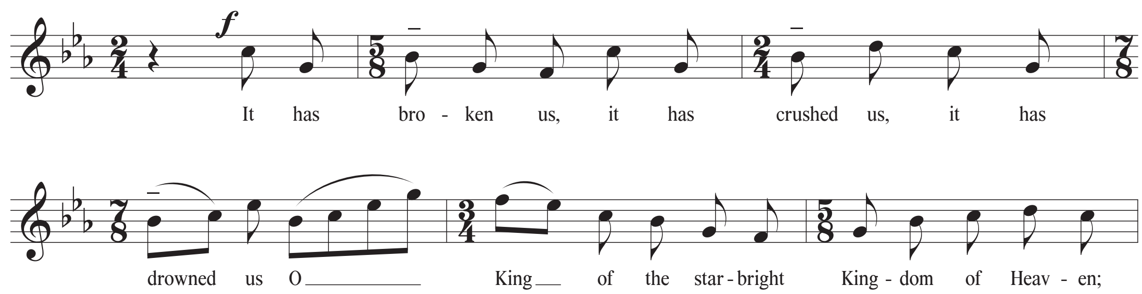
Devoured appears four times in this analysis as potentially difficult, and *starbright* appears three. *Crimson, has, us, King, of, Heaven, consumed, timber, drowned,* and *swallowed* appear twice. Words that appear once are *Kingdom*, *wind, broken, crushed, from, O, wind, as* and *by.* The monosyllables are *has, us, of, King,* and *drowned.* We can pay particular attention to these words in the analysis of the musical setting.

Analysis of the musical setting is as follows.

Metric setting

"Sea-Snatch" is set primarily in alternating measures of 5/8 and 2/4, with occasional appearances of 7/8, 6/8 and 3/4.[138] The metric stresses align well with the sentence and word stresses throughout; indeed, the meters seem to have been chosen in many cases specifically for this purpose. (See Example 9.1.) This should contribute to overall intelligibility; however, see the section on high suprisal words below.

Example 9.1. Measures 4-9, "Sea-Snatch" from *Hermit Songs* by Samuel Barber (vocal line only). Time signatures added.

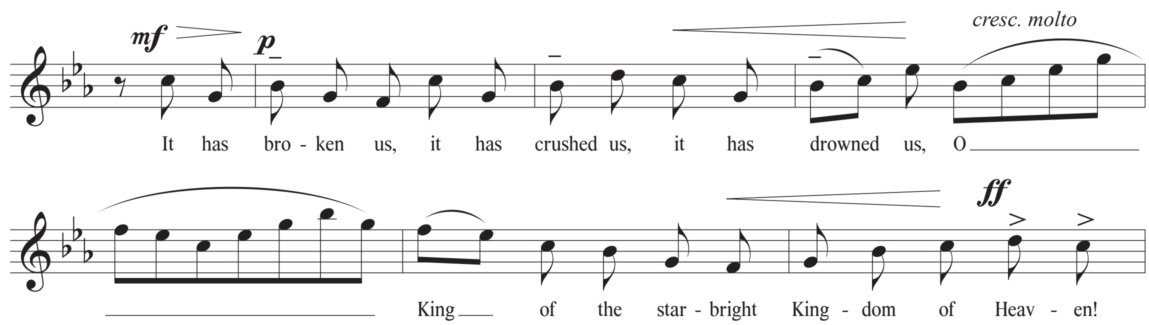


Word Groups

The integrity of word groups is maintained throughout the setting, with the possible exception of the second setting of the line "O King of the starbright Kingdom of Heaven!" There, the long melisma on "O" could have the effect of making the word groupings ambiguous (although the fact that the words are being repeated here may mitigate that fact). (See Example 9.2.) There is further discussion of this melisma later in this analysis.

[138] There are no time signatures in the score, but beat groupings are made clear by beaming.

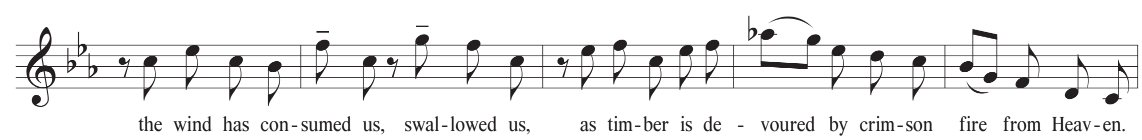
Example 9.2. Measures 24-30, “Sea-Snatch” from *Hermit Songs* by Samuel Barber (vocal line only).



Tessitura

In the original (higher) key, the piece stays within the treble staff except for during the two melismas on “O” and in three other places. Those other spots are settings of the “difficult” words *consumed, swallowed* and *devoured.* (See Example 9.3.) All three of those words are bisyllabic, which makes them somewhat less vulnerable to mishearing; but *devoured* is the most-flagged word in the text analysis. It is quite likely that *devoured,* and somewhat likely that the other two words will also will not be intelligible.

Example 9.3. Measures 14-18, “Sea-Snatch” from *Hermit Songs* by Samuel Barber (vocal line only).



High Surprisal Words

Devoured, crimson, and *starbright* are uncommon enough words to be high surprisal in many contexts; as noted above, there is high surprisal throughout the text. In speech, high surprisal words are given more time than low surprisal words, and this setting, in a fast tempo, in which no syllable is given more than a quarter note and most get an eighth note, does not attempt to duplicate that at all. No metronome marking is specified; the tempo marking specified is “Allegro con fuoco, surging.” Leontyne Price’s recording makes the quarter note about 156, which works out to a rate of 312 syllables per minute, or 5.2 syllables per second. Figures for the average rate of English speech in syllables per second vary from 3.5 to 6.2;[139] certainly this pace falls within that range, although at the upper end. However, the absence of any longer setting for high surprisal words is all the more telling in this otherwise speechlike rhythmic context.

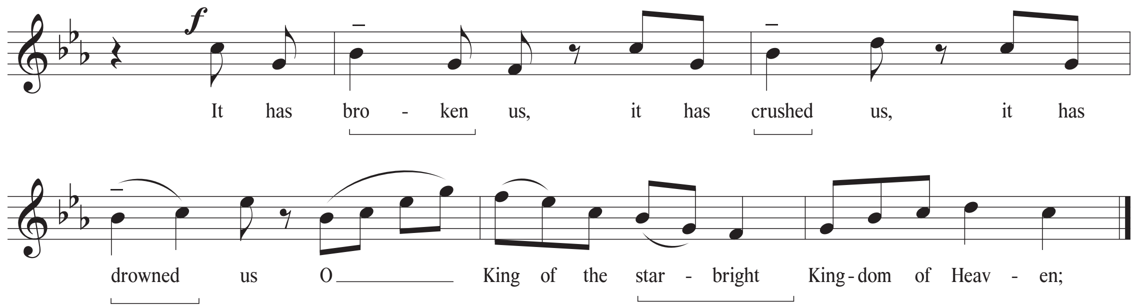
Two counter-examples will illustrate. In the first, I rewrite the text for the first phrase using lower-surprisal words. (Example 9.4.) In place of *broken, crushed* and *drowned,* I have *substituted given, made* and *taken*—the three words that are listed by the Google Ngram Viewer as the most frequent words between *it has* and *us.* In place of *starbright,* I have substituted *royal* to modify Kingdom.

Example 9.4. Measures 4-9 of “Sea-Snatch” (vocal line) replacing high surprisal words with low surprisal words.

[139] François Pellegrino, J. Farinas and J.-L. Rouas, “Automatic Estimation of Speaking Rate in Multilingual Spontaneous Speech,” *ISCA Archive, Speech Prosody 2004, Nara, Japan*: 2 http://ronnetsell.net/html/rate_of_speech_and_comms_effic.html, (accessed 5/20/17); François Pellegrino, Christophe Coupé, Egidio Marsico, “Across-Language Perspective on Speech Information Rate,” *Language* 87 (September 2011): 544.

In the second, I use the original text and pitches but alter Barber's rhythm to give the high surprisal words more time. (Example 9.5.)

Example 9.5. Measures 4-9 of "Sea Snatch" (vocal line) with the rhythmic setting altered to give high surprisal words more time.



In my own audiation of these alternatives, it seems clear to me that either would be easier to comprehend than the original.[140] This is, of course, subjective.

Perception of vowel substitution

Because there are no held vowels in this setting, this effect should not be a major contributor, though it is possible that *King* will be heard as *keng* and/or that *Heaven* will be heard as *havven*. (See Example 9.2.)

Unvoiced fricatives

As noted above, *swallowed* is vulnerable to mishearing because of its initial consonant.

[140]I am not, of course, suggesting that Barber should have set the text differently or somehow "better," merely illustrating what effect his compositional choices may have had on the intelligibility of his text.

Chunk Density

At ♩ = 156, the two-second limit is reached in just over two bars of the initial alternation of 5/8 and 2/4. In that time, Barber fits a repetitive alternation of two right-hand chords over the running eighth note pattern, which occurs twice at the beginning of the piece before the vocal enters over it. This is essentially the same strategy as the musical theatre vamp, and the pattern should be easily understood as a chunk by the time the listener is also processing words. The pattern changes in bar seven, but the material combines the steady stream of eighth notes that has been ongoing in the left hand and the melodic pattern of ascending perfect fourths that the right hand had been outlining. While it is a different texture, it seems likely that it will be perceived as one large chunk rather than a stream of new information, and intelligibility will not suffer further thereby. (See Example 9.6.)

Example 9.6. Measures 1-9 of "Sea-Snatch" from *Hermit Songs* by Samuel Barber. Accompaniment chunks in brackets (added).

However, the 11-note melisma on the word *O* in measures 27-28 (see Example 9.2) may or may not be perceived as chunks—while it does stay in the pentatonic minor scale that has made up the rest of the phrase, it contains a melodic contour not seen hitherto in the song, which would work against chunkification (if I may be permitted to coin a word).

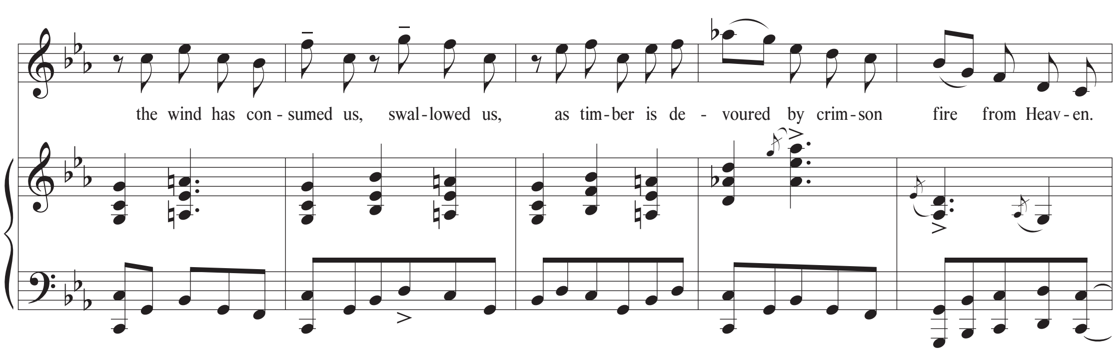
Recognition points

At this tempo, holding words in working memory until their recognition points is not going to be an issue. The only exception is the same melisma on the word *O* just discussed. Since the recognition point of *O* is the beginning of the subsequent word, and the melisma lasts just over two seconds at Leontyne Price's tempo, it is possible that the listener will have lost track of the beginning of the word by the end of it. As it is a monosyllable, this would seem to be of little import; but as we have identified this both as a potential information overload and as a difficult sentence structure as well, it seems likely that this could be another spot in which intelligibility is especially compromised. Working against this, of course, is the fact that this is a repetition of text—assuming it was understood the first time.

Harmonic priming

As the song is harmonically static, being essentially an elaboration of the C minor pentatonic scale, there is virtually no harmonic motion and thus no effect on intelligibility from harmonic priming or the lack of it. The one significant harmonic effect is a change from A-natural to A-flat on the 2nd syllable of the word *devoured* (see example 9.7), which indeed is harmonically unexpected, even over the continuation of the left-hand ostinato; however, this as this word is likely to be unintelligible for multiple reasons already, this is simply icing on the cake.

Example 9.7. Measures 14-18 of "Sea-Snatch" from *Hermit Songs* by Samuel Barber.



Repetition

It is possible that the final phrase of the song will be more intelligible because it is a repetition of the first phrase, despite the intervening material.

Conclusions

Many aspects of Barber's setting conspire to make the text of this song extremely hard to understand. The sentence structures of the second line (a long direct address) and the fourth (a long sequence of stacked prepositions) may increase the neural processing cost, as may the constant switching of sentence structures. The same is true of the cavalier treatment of high surprisal words; as my counterexamples showed, such a rapidly rhythmic delivery may be intelligible with low surprisal words, but in order for such a high surprisal text to be intelligible, the rhythmic setting would have needed to take the surprisal into account. The "difficult" words *devoured, consumed,* and *swallowed* are set at the top of the staff, where their intelligibility will suffer accordingly; *devoured* is potentially done even more damage by its harmonic setting. The long melisma on the monosyllable *O* potentially runs afoul both of recognition point and chunk density factors. In addition, several words of the text also have a large potential for phonetic ambiguity, with easily made substitutions among the various families of consonants contributing to possible mishearings of the text.

The song has an undeniable dramatic appeal, but a soprano who cannot make this song's text intelligible need not blame herself.

Chapter 10: Analysis, "Hymn."

"Hymn" from *Serenade for Tenor, Horn and Strings* (first stanza) by Benjamin Britten (on a text by Ben Jonson).

This song cycle was composed in 1943 for hornist Dennis Brain and tenor Peter Pears, Britten's partner and muse. It has become a staple of the tenor repertoire, having been performed and recorded by countless singers; the Naxos music library has recordings featuring fifteen different tenors.[141]

I include an analysis of this piece here both to include music written for tenor and to include a piece in which classical poetry is used as text.

The text of the first stanza:

Queen and huntress, chaste and fair,
Now the sun is laid to sleep,
Seated in thy silver chair,
State in wonted manner keep:
Hesperus entreats thy light,
Hesperus entreats thy light,
Goddess, goddess, goddess, excellently bright.[142]

Analysis of this text using the methods outlined in Chapter 6 yields the following.

---

[141] https://imslp.naxosmusiclibrary.com/work.asp?wid=94266. Accessed 5/23/17.
[142] The song continues:

Earth, let not thy envious shade
Dare itself to interpose;
Cynthia's shining orb was made
Heav'n to clear when day did close:
Bless us then with wishèd sight
Goddess, goddess, goddess excellently bright.
Lay thy bow of pearl apart,
And thy crystal shining quiver;
Give unto the flying hart
Space to breathe, how short soever:
Thou that makst a day of night,
Thou that makst a day of night,
Thou goddess, goddess, goddess, goddess, goddess excellently bright.

Sentence Structure

Like much classical poetry, the sentence structure here is convoluted. The first line (*Queen and huntress, chaste and fair*) is a direct address, much like the beginning of "Green Finch and Linnet Bird," although the addition of the adjectives complicates things somewhat; there may be a slight processing cost here. The next line *(Now the sun is laid to sleep*) is a relative clause, although the omission of the word *that* after *now* makes it impossible to tell this until after the rest of the sentence; in practice, this will probably be heard as a sentence in its own right, with no additional difficulty. The third and fourth lines are inverted from normal English sentence order, which would be this: *Keep state in [thy] wonted manner, seated in thy silver chair.* Comprehension will almost certainly be compromised here, even if intelligibility is not.

Hesperus entreats thy light is a simple sentence, although it then has another direct address tacked onto it: *goddess, goddess, excellently bright.*[143] This should not be a major impediment to intelligibility.

Surprisal

Chaste, thy, wonted, and *entreat* do not appear on the list of the 20,000 most common English words, and are potentially difficult. While neither does *huntress*, and *goddess* comes in at 6,737, their cognates *hunter* and *god* are both among the 5,000 most common words; these two should pose no problem.

[143] The repetition of the word *goddess* is Britten's.

Hesperus[144], as a proper noun, is of course not on these lists. While it would likely have been more familiar to Britten's mid-20th century British audience, today this is an obscure reference, and will be difficult to process.

Interestingly, there are no other major contributors to high surprisal. *Huntress* as a continuation of *queen and* is only 100 times less likely than the most frequently occuring word, *her*, according to the Google Ngram Viewer. And while *chair* is not a frequent follower of the word *silver*, as a continuation of *Seated in thy silver* it is nearly mandatory.

<u>Cohort size</u>

The following words have cohort sizes of twelve or more common words after the first vowel (except as noted): *and, huntress, fair, sun, to, seated, silver, state, manner,* Hesperus*, entreats, goddess,* and *excellently.* Of these, only *fair* and *excellently* are set to long enough note values that their recognition points might come into play. Please see the Appendix for discussions of the cohorts and recognition points of each word.

<u>Phonetic ambiguity</u>

•The only words containing /b/, /d/, or /g/ are *laid, goddess* and *bright.* Of those, the only possible substitution that creates another word is /b/ for /g/ as the initial sound of *goddess,* creating *bodice*, and then only if the second syllable is pronounced as a schwa. This seems unlikely to create much difficulty.

---

[144] The evening star; a term from Greek mythology.

•*Sleep* could be misheard as *sleek* or *sleet; state* could be misheard as *skate*, *spate* or *stake/steak*; *keep* could be misheard as *peep*; *light* could be misheard as *like. Keep* is the most vulnerable, as the substitution is in the initial position.

•There are several words beginning with unvoiced fricatives: *huntress, fair, sun, sleep, seated, silver* and *state* (not to mention *Hesperus*). However, these are unlikely to be a problem, as the whole passage is set quite softly.

•Words beginning with consonants and followed by front vowels include *queen, chaste, fair, laid, sleep, seated, chair, state, manner, keep,* and *Hesperus.*

•Words containing a back vowel followed by a consonant include only *goddess.*

Of the words identified above, the ones that would seem to be at most risk are *sleep* and *state,* which are vulnerable for two reasons and are monosyllables to boot, and *goddess,* which likewise is identified twice as potentially problematic.

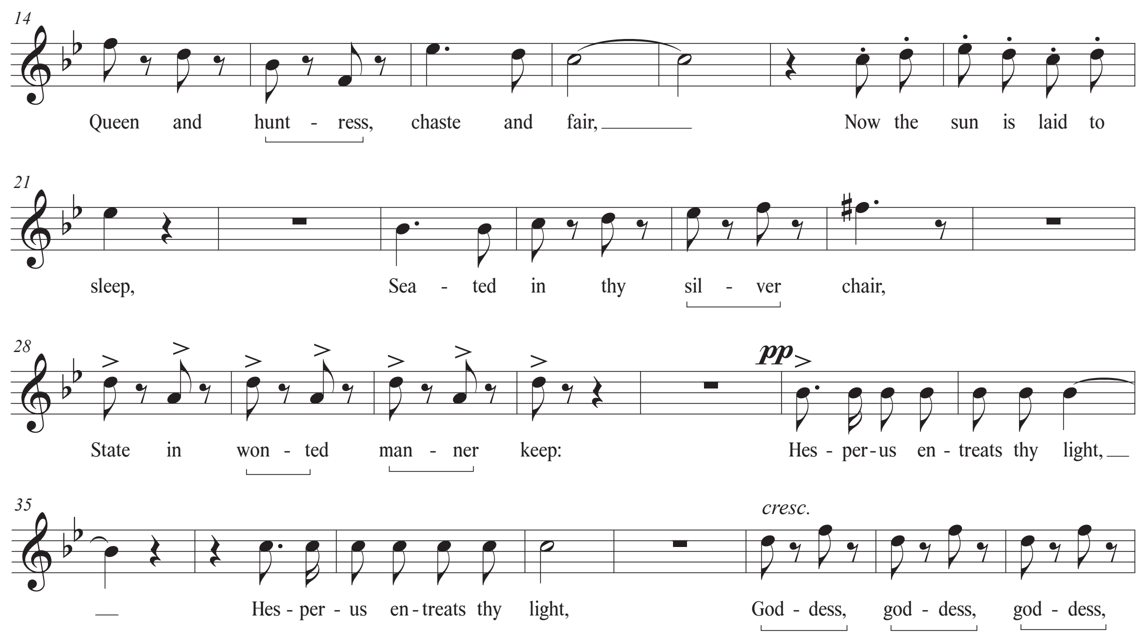
"Difficult" words

Goddess arises as potentially problematic in this analysis four times. *State* and *fair* appear three times. Words that appear twice are *chaste, entreats, huntress, Hesperus, excellently, seated, manner, sleep,* and *keep.* Words that appear once are *thy, wonted, and, sun, to, silver, light, queen, laid* and *chair.* Any of these words that are given problematic settings will be noted in the musical analysis below.

Metric setting

The only conflict between the poetry's meter and the metric setting is the second statement of the word *Hesperus,* which is set with the final (normally unstressed) syllable on the downbeat. This would certainly degrade the intelligibility of an already obscure word, unless there is a mitigating influence of the repetition. The only exceptional feature of the rhythmic setting is the staccato setting of several words, on eighth notes alternating with eighth rests. (See Example 10.1.)

Example 10.1. "Hymn" m14-42 (vocal line only) from *Serenade for Tenor, Horn and Strings* by Benjamin Britten. Two syllable words with eighth rests between the syllables are marked with brackets.



Particularly when these rests separate syllables of a bisyllabic word, the potential exists for them to interfere with intelligibility. This is lessened by the extremely quick tempo (♩ = 168 - 170), but only somewhat. At that speed, a two-syllable word set this way will take up about ¾ of a second; recall that the average number of syllables per second in spoken English is between 3.5 and 5.2. So this is still a slow pace compared to spoken English. The separation of syllables can only confuse the cohort-narrowing process, if observed literally.[145]

Word Groups

The integrity of word groups is maintained throughout the setting.

Tessitura

As this excerpt stays safely within the normal tenor range, tessitura is not a factor in intelligibility.

High Surprisal Words

The phrase *Queen and huntress, chaste and fair* contains two high surprisal words: *huntress* and *chaste.* Helping with their intelligibility, they are indeed set with longer rhythmic values than the other words in the phrase (except *fair,* which is sustained for four beats). *Thy* is one of the shorter words in its phrase, *seated in thy silver chair*, which could detract from its intelligibility. *Wonted* is longer than *State, in* or *keep,* but the same length as the much more common *manner.* There may be some small contribution to unintelligibility there.

Interestingly, Britten sets *Hesperus entreats thy light,* which contains three high-surprisal words back-to-back, essentially at double the tempo of the previous two sentences, which would seem to be an intelligibility disaster. All three of them are shorter

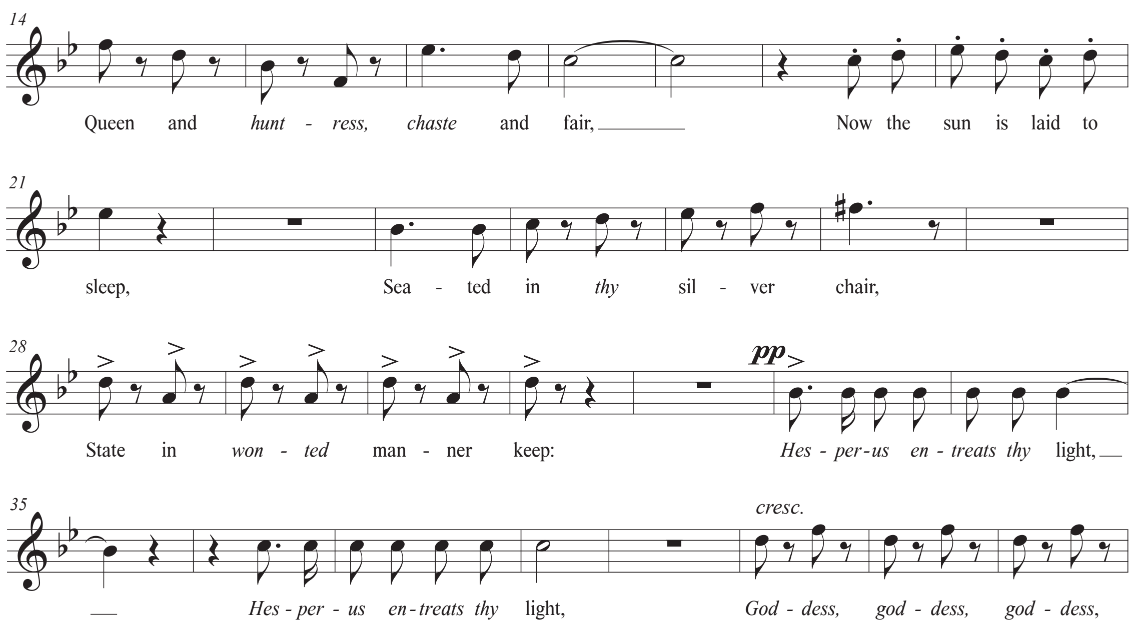
---

[145] In Peter Pears' recording, as one example, the rests seem to have been taken simply as a reinforcement of the *leggiero* instruction; the rests between syllables are not of equal length with the syllables themselves.

than the much more common word *light.* The line is repeated immediately, which may help with intelligibility (Huron and Collister's research weighed in on the effect of immediate repetition of words, not phrases), although, as previously noted, the second appearance of the word *Hesperus* is a metric mismatch.

The word *Goddess* is immediately repeated, taking advantage of the increased intelligibility of repeated words, and is given a longer duration in the rhythmic setting than most of the words thus far; it should pose little difficulty. (See Example 10.2.)

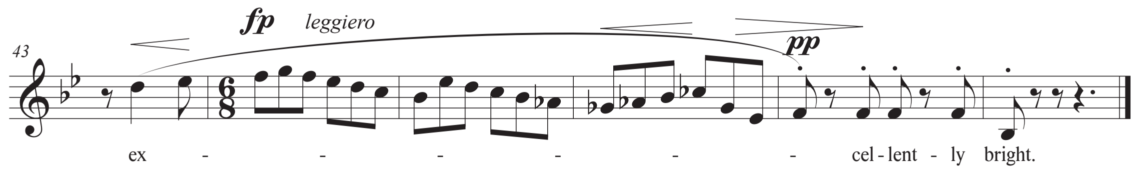
Example 10.2. "Hymn" m14-42 (vocal line only) from *Serenade for Tenor, Horn and Strings* by Benjamin Britten. High surprisal words are in italics.



Perception of vowel substitution

The first vowel of the word *excellently* rises above the threshold at which it will start to be heard as /æ/ for three notes of its 21-note melisma. (See Example 10.3.) This will certainly add to the uncertainty surrounding the melisma (see below), but will probably not be a decisive factor.

Example 10.3. "Hymn" m43-48 (vocal line only) from *Serenade for Tenor, Horn and Strings* by Benjamin Britten.



Chunk Density

The song begins with a thirteen measure introduction for the horn and strings; the horn part of this is then repeated nearly exactly as the accompaniment for the first four lines of the poem, with a less active part for the strings. The horn is quite active, and could easily increase chunk density to the level of processing difficulty in certain spots. However, the fact that the horn part will have been heard once before the necessity arises of processing it in parallel with the vocal line should help somewhat in those spots in which it is indeed repeated exactly.

The most often cited as "difficult" words in this text are *fair* and *state. Fair* coincides with the beginning of the restatement of the horn part, although it has been modified a bit. (See Example 10.4.) While it's likely that the slightly different form of the horn part in measures 16 and 17 will prevent it from being recognized, and thus increase chunk density, the horn part in measure 18 is a distinctive melody that should be processable as one chunk. *Fair*'s intelligibility will likely suffer a bit from the increased chunk density at the beginning of the word, but perhaps will recover by the end.

Example 10.4. "Hymn" m1-2 and 16-18 from *Serenade for Tenor, Horn and Strings* by Benjamin Britten.

The word *State,* on the other hand, will not suffer at all, as the horn part underneath it is the beginning of the third repetition of a pattern in the horn part. (See Example 10.5.)

Example 10.5. "Hymn" m26-28 from *Serenade for Tenor, Horn and Strings* by Benjamin Britten.

The line *Hesperus entreats thy light* is accompanied by a new pattern in the strings. It is repetitive, and should be recognizable as a chunk or chunks during the second statement of the line, but the appearance of a new, rhythmically active accompaniment pattern during the line will definitely add to the difficulty of understanding an already difficult setting. However, the pattern repeats through the triple setting of the word *goddess,* and will not be at all distracting by that point. (See Example 10.6).

Example 10.6. "Hymn" m32-42 from *Serenade for Tenor, Horn and Strings* by Benjamin Britten.

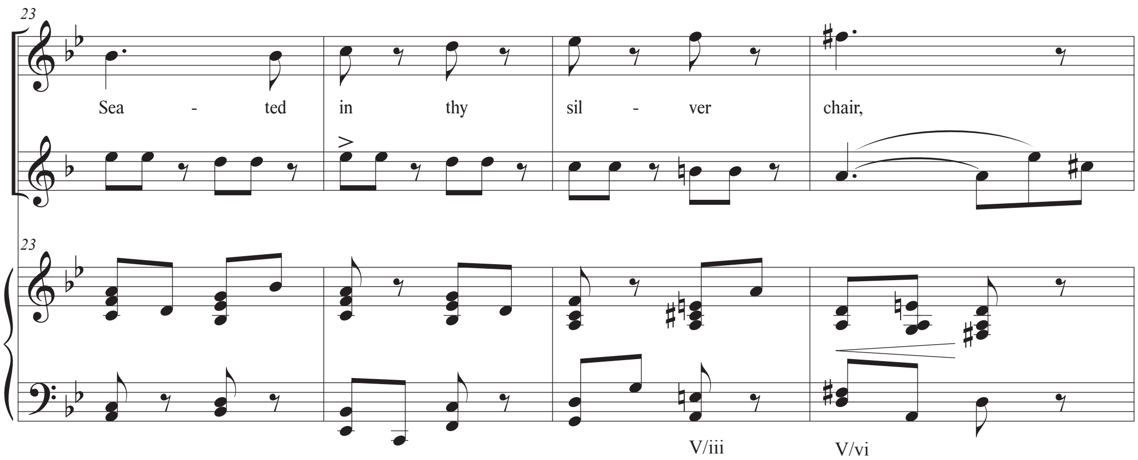
Recognition points

The only word held more than the two-second limit of working memory is *excellently.* (See Example 10.3, above.) This case, where the first syllable of a polysyllabic word is set melismatically and the subsequent syllables are set syllabically, is an ingenious example of an end run around the difficult aspects of both melismas and long notes for intelligibility. While a listener will have absolutely no idea what the word is for the entirety of the melisma, other than that it starts with the vowel /ɛ/, by the time she hears the first /l/, the cohort will have narrowed down to *excellent* and its cognates, and there will be no doubt about the identity of the word at its end.

Harmonic priming

Because the harmonic structure of half of the first stanza is laid out first in the introduction, it's debatable whether anything in that section could be dubbed harmonically unexpected. The best candidate for that honor if we ignore the introduction would be the setting of the word *chair,* which occurs as the harmony changes to a bracing V/vi after a rapidly changing series of harmonically ambiguous chords seemingly determined as much by voice leading considerations as harmonic ones (the last of which, however, is V/iii—the dominant of the V/vi chord). (See Example 10.7.) As *chair* is not otherwise a particularly difficult word, though, this is unlikely to have a major effect.

Example 10.7. “Hymn” m23-26 from *Serenade for Tenor, Horn and Strings* by Benjamin Britten.



Conclusions

Certain words’ or phrases’ intelligibility could suffer in this setting. The inversions of normal sentence order will impose a processing cost. Not all of the high surprisal words are set with longer rhythmic values than the other words in their sentences. Mental processing of *fair* is liable to be disrupted by the chunk density of the horn part. The double time setting of the line *Hesperus entreats thy light*, particularly, will not make the string of high surprisal words any easier to decode—especially since any mitigating effect of the line’s repetition is liable to be unmitigated by the metric mismatch of the repetition.

However, Britten has successfully exploited some of the characteristics of language processing we have been discussing to make certain words more intelligible in their settings. The word *goddess,* potentially difficult, is repeated. The word chosen for an extended melisma, *excellently*, is polysyllabic and thus has a high effective frequency; and the ambiguity of the melisma is instantly resolved by the syllabic setting of all of the word after the first vowel.

Chapter 11: Analysis, "Five Fathoms Deep."

"Five Fathoms Deep" from *The Tempest* by Thomas Adès (on a text by Meredith Oakes).

The Tempest was premiered at the Royal Opera House in 2004, receiving a warm reception from critics, and has subsequently been seen in (as of this writing) eight other cities. It is widely regarded as the high point of the composer's career to date.

Ariel's aria "Five Fathoms Deep" is renowned both for its inventive and striking writing for the coloratura soprano—and for the sheer impossibility of understanding any of the words.

The text is as follows:

Five fathoms deep
Your father lies
Those are pearls
That were his eyes
Nothing of him
That was mortal
Is the same
His bones are coral
He has suffered
A sea change
Into something
Rich and strange

Sea nymphs hourly
Ring his knell
I can hear them
Ding dong bell

Analysis of this text using the methods outlined in Chapter 6 yields the following.[146]

Sentence Structure

There is no punctuation in the text, but it is not difficult to sort it into sentences, as follows:

1) Five fathoms deep your father lies.
2) Those are pearls that were his eyes.
3) Nothing of him that was mortal is the same.
4) His bones are coral.
5) He has suffered a sea change into something rich and strange.
6) Sea nymphs hourly ring his knell.
7) I can hear them, ding dong bell!

Sentences 1 and 2 feature inverted sentence order (compare "Your father lies five fathoms deep" and "Those that were his eyes are pearls"). In sentence 6, the position of *hourly* before the verb is unusual (compare "Sea nymphs ring his knell hourly"). These deviations from more usual sentence structure will likely come at a cost to intelligibility.

Surprisal

Words that do not appear on the list of the 20,000 most common English words are *fathoms* (although *fathom* is number 13,898 as a verb), *knell, ding,* and *dong*. However, *dong* is the most common word after *ding*, according to the Google Ngram viewer; accordingly, I will not count it as high-surprisal in this context. Other less

[146] As this is a song that Ariel sings to Ferdinand, not entirely connected to the plot, there is little dramatic context to aid comprehension here. The fact that the text is similar to, but condensed from, Shakespeare's version may both hurt and hinder the recognition of words for those familiar with the original.

common words include *pearls* (7,336), *mortal* (9,747), *coral* (6,124), *nymphs* (16,155), and *hourly* (12,255). Depending on the setting, these words may be difficult to decode.

The only word that seems unusual enough in its context to be high-surprisal without reference to frequency is *coral*, although it will be helped somewhat by the context (assuming that *fathoms* has been understood). It is worth noting that *coral* will likely not get a predictability boost from being the second member of a rhyming pair, as it is not a perfect rhyme with *mortal.*

<u>Cohort Size</u>

The following words have cohort sizes of twelve or more common words after the first vowel (except as noted): *fathom, deep, lies, pearls, his, of, him, mortal, same, coral, suffered, something, and, hourly, can, hear, ding, bell.* Please see the Appendix for discussions of the cohorts and recognition points of each word.

<u>Phonetic ambiguity</u>

•Words containing /b/, /d/, or /g/ are *deep, bones, suffered, and, ding, dong,* and *bell.* Possible substitutions that create words are *beep* for *deep, gong* for *dong,* and *dell* for *bell*; these words have an associated unintelligibility risk.

•Words containing /p/, /t/, or /k/ are *pearls, that, mortal, coral, into,* and *strange.* The only possible substitution that creates a word is *curls* for *pearls.* The risk is similar.

•Words beginning with unvoiced fricatives include *five, fathoms, father, that, his, him, he, has, suffered, sea, strange,* and *hear.* Of these, *fathoms* is set ***ff*** with its first syllable on E6; *father* is set ***mf*** with its first syllable on A♭5. These words' intelligibility will definitely suffer. Words set on high notes but at a dynamic of ***p*** or softer include *his* (B5), *him* (F#5), *has* (F#5), *sea* (B5 and F#5), *something* (first syllable on B5), and *strange* (E6); these words' intelligibility may be affected.

• Words beginning with consonants and followed by front vowels include *fathoms, deep, that, same, has, sea, change, strange, knell, can, them,* and *bell.*

•Words containing a back vowel followed by a consonant include *your, father, mortal, coral,* and *dong.*

Of these, it seems that *fathom, father, has* and *see* are most at risk for unintelligibility.

“Difficult” words

Fathoms appears four times in this analysis as potentially problematic. Words that appear three times are *pearls*, *mortal*, *coral*, *deep, father* and *bell.* Words that appear twice are *hourly, knell, ding, dong, his, him, same*, *can, has, sea* and *strange.* Words that appear once are *lies, of, suffered, something, and, hear, him, something, that, same, change, them, nymphs,* and *your.* Any of these words that are given especially problematic settings will be noted in the musical analysis below.

Metric setting

Adès has set the the text with each syllable receiving the same rhythmic value: a dotted half note, with the half note between 40 and 44 beats per minute. (The words “I can hear them” are the only exception; these receive a dotted eighth note each at the same tempo.) As the orchestra plays in rhythmic unison with the vocal line for the majority of the piece (and does not set up a strong pulse in the passage where it has other material), the effect is that of each syllable getting one slow pulse at between 27 and 29 beats per minute.

The sense of meter is thus extremely attenuated, but accented syllables in the text are indeed lined up with downbeats. (See Example 11.1.)

Example 11.1. “Five fathoms deep” (vocal line only) from *The Tempest* by Thomas Adès. Mis-accented syllables are underlined.

It is worth noting, however that there are a few words in the text that are placed in the poetic meter so as to be mis-accented, whose accents Adès observes just as meticulously; these, of course, may be harder to understand. These syllables are underlined in Example 11.1.

Because the sense of meter is so faint, other forms of accent take on an unusual significance in this regard. Specifically, the leap of a major 9th to E6 between the 1st and 2nd syllables (along with the dynamic swell to ***ff*** and the orchestra's ***fff*** chord) is what establishes that there is a downbeat at all. Conversely, the other ***ff*** E6 in the vocal line, on the word *are* (in the 3rd system, 4th bar of Example 11.1), thoroughly disrupts what sense of meter there is, coming as it does on an offbeat and on a normally unstressed word. This will create an intelligibility cost as well.

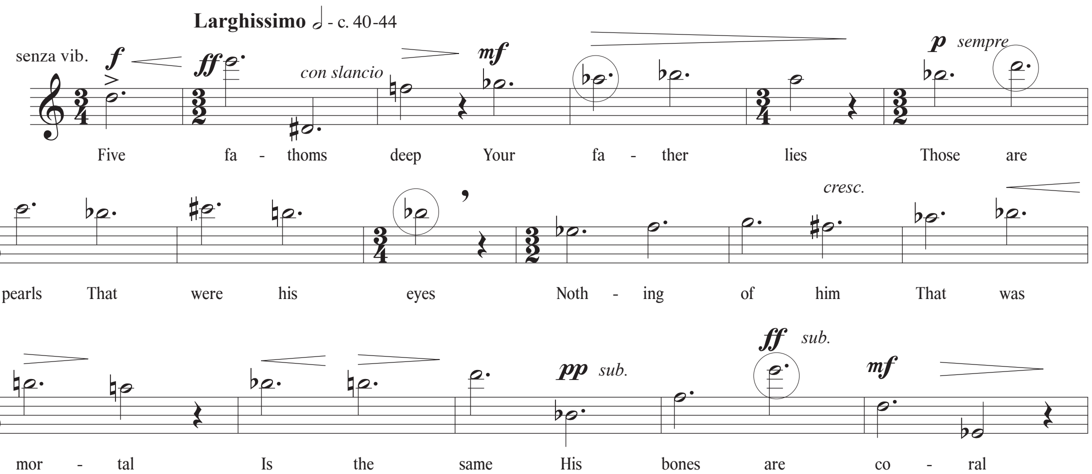
Word Groups

The integrity of word groups is maintained throughout the setting, at least as far as rhythmic setting is concerned. However, the extreme leaps in the setting of "five fathoms deep" may have the effect of separating the syllables (and thus the words) from each other in the listener's ear. The same is probably true of "ding dong bell." There may be an associated cost to intelligibility.

Tessitura

There are fifty-nine pitches in the vocal line of this aria. Forty-nine of these are at F5 (the pitch at which all vowels begin to sound like /ɑ/) or above. Of those 49, only four actually are syllables containing /ɑ/; the other 45 will most likely have their vowels misperceived (see Example 11.2).

Example 11.2. “Five fathoms deep” (vocal line only, 1st 17 measures) from *The Tempest* by Thomas Adès. Syllables above F5 whose vowels are /ɑ/ have their notes circled; these may have their vowels perceived correctly. All other words set above F5 will likely have their vowels misperceived.



High Surprisal Words

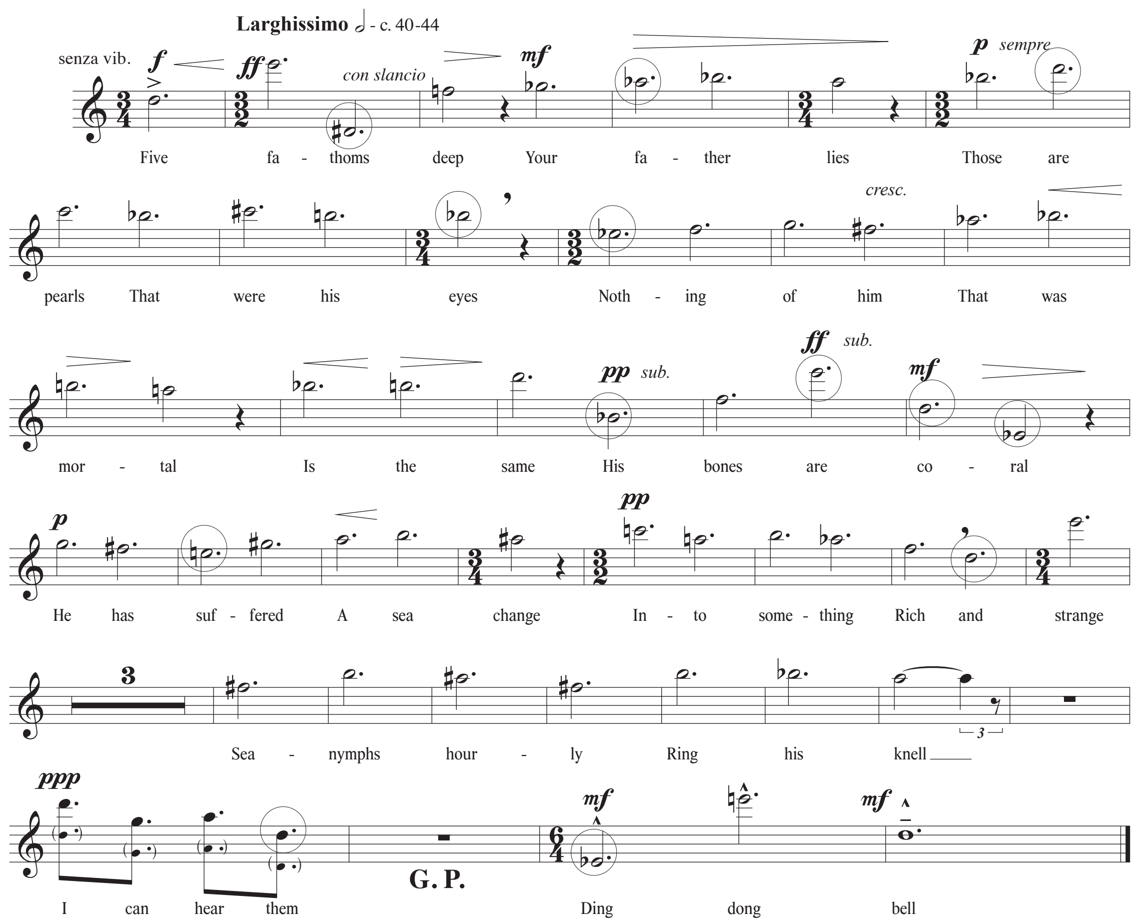
The high surprisal words in this text are *fathoms, pearls*, *mortal*, *coral*, *nymphs*, *hourly*, *knell, ding,* and *dong. Coral* and *ding* are the only ones set entirely below F5; they are the only ones that are at all likely to be intelligible.

Because the rate of delivery of the words of this aria is so much slower than the rate of delivery of spoken English, the question of whether high surprisal words get extra time is moot.

Perception of vowel substitution

As mentioned in Chapter 6, it is primarily in words or syllables that are held out long enough for the impact of formant transitions to be diminished that this effect is most noticeable. Nearly all of this aria qualifies. We have already established that the portions of this aria set above the staff are likely to be have their vowels misperceived; what about the others? If we examine the ten syllables on notes that are not on F5 or above we discover that only the final word, *bell*, is likely to fall prey to this effect. Taking these lower-set notes together with the four syllables set above the staff whose vowel is /ɑ/, Example 11.3 shows the syllables in the aria whose vowels are likely to be perceived correctly. There are only thirteen of them.

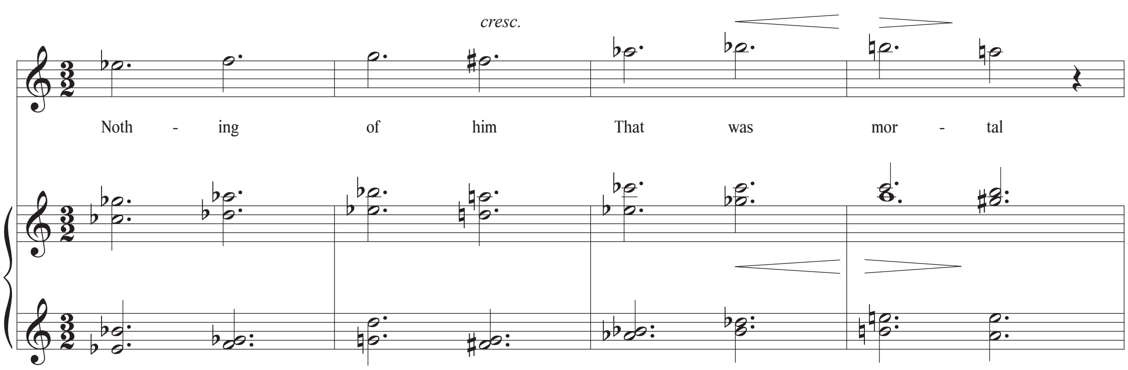
Example 11.3. "Five fathoms deep" (vocal line only,) from *The Tempest* by Thomas Adès. Circled notes are the only ones likely to have their vowels perceived correctly.



Chunk Density

For most of the aria the orchestra moves in rhythmic unison with the vocal line; there is no interference with intelligibility from the accompaniment. Example 11.4 shows a typical passage.

Example 11.4. “Five fathoms deep” from *The Tempest* by Thomas Adès. First four measures of rehearsal 91.



The only change in texture comes under the line “Sea nymphs hourly ring his knell” (see example 11.5). At this tempo, however, even this comparatively active accompaniment does not risk overloading the chunk capacity of working memory. Moreover, it is introduced four bars earlier, so that the couple of events per measure that occur here may indeed be perceived as belonging to larger chunks.

Example 11.5. “Five fathoms deep” from *The Tempest* by Thomas Adès. First four measures of rehearsal 92

Recognition points

Even if taken at the upper end of the tempo range given, each syllable of the vocal line (excepting “I can hear them”) will last for more than two seconds, the upper temporal limit of working memory. It is thus going to take active concentration (“rehearsal,” in the cognitive science sense) for the listener to retain the beginnings of syllables until those syllables end. Even if the beginning of a word is successfully kept in mind, the length of the syllables means that the cohort of each word will remain unresolved for at least two seconds. Given the distortion of most of the vowels due to the high tessitura and the possible ambiguities detailed above, the cohorts may include many spurious candidates, possibly to the exclusion of the actual words in the text.

Moreover, the bisyllabic words in the text will not receive their usual advantage in intelligibility over monosyllables, since their recognition points, even if at the earliest possible point in the word (the first consonant of the 2nd syllable), will be more than two seconds separated from the beginning of the word. All of this seems likely to contribute greatly to the difficulty of deciphering the text in this setting.

Harmonic priming

Given the richness, variety and unpredictability of Adès' harmonic language, and the lack of any hierarchical harmonic relationships on display in this aria, it is difficult to argue for any significant effect of harmonic priming on intelligibility.

Conclusions

Everything about this setting works together to make the text almost entirely unintelligible. The unusually slow tempo makes active rehearsal necessary to decode even a monosyllable. The unusually high tessitura makes the proportion of vowels that are likely to be perceived correctly less than a quarter of the total, making that active rehearsal even more difficult.

While these are the major contributors, there are other factors that point in the same direction. Roughly one-sixth of the words in the text begin with unvoiced fricatives and are set either high or high and loud, thus vulnerable to mishearing. Large leaps may disrupt the word groups *five fathoms deep* and *ding dong bell*, both of which also contain relatively high surprisal words, making them harder to process. There are normally unaccented words in accented positions in the poetic meter, which the setting faithfully replicates, at a further cost to intelligibility. A major dynamic stress on a very high note on a normally unstressed word (*are*) on an offbeat disrupts what tenuous sense of meter there is, potentially making matters worse.

Intelligibility was presumably not one of Adès' major or even minor goals for this aria. In constructing it to other aesthetic purposes, he has given us the opportunity for a telling analysis of how to make the unintelligibility of an English text setting proof against a singer's best efforts.

Chapter 12: Analysis, "Some Enchanted Evening."

"Some Enchanted Evening" from *South Pacific* by Richard Rodgers (music) and Oscar Hammerstein II (lyrics).

The Broadway musical *South Pacific* opened in 1949, and has become one of the pillars of the American musical theatre canon, one of Rodgers and Hammerstein's most beloved shows, the winner of multiple Tony awards and the Pulitzer Prize for drama.[147] "Some Enchanted Evening" is one of many hit songs from the musical, and one of those sung by the character of Emile de Becque, a role originally written for the operatic bass Ezio Pinza, who had turned to Broadway and popular entertainment after his opera career.[148]

As the song was written to be sung over an orchestra by an operatic voice without amplification, it bears direct comparison with the opera and art song literature for our purposes in a way that "Green Finch and Linnet Bird" perhaps does not, as *Sweeney Todd* was first produced in the era of Broadway amplification.

The text of the song is as follows:

Some enchanted evening you may see a stranger,
You may see a stranger across a crowded room
And somehow you know, you know even then
That somewhere you'll see her again and again.

Some enchanted evening someone may be laughing,
You may hear her laughing across a crowded room

[147] *Encyclopedia of Popular Music*, 4th ed., "South Pacific (stage musical)," Oxford University Press, http://www.oxfordmusiconline.com.proxy.libraries.rutgers.edu/subscriber/article/epm/62614 (accessed August 18, 2017).
[148] *Grove Music Online*, "Pinza, Ezio" Oxford University Press http://www.oxfordmusiconline.com.proxy.libraries.rutgers.edu/subscriber/article/grove/music/21795, (accessed August 18, 2017).

And night after night, as strange as it seems
The sound of her laughter will sing in your dreams.
Who can explain it who can tell you why?
Fools give you reasons, wise men never try.

Some enchanted evening when you find your true love,
When you feel her call you across a crowded room,
Then fly to her side and make her your own,
Or all through your life you may dream all alone.

Once you have found her, never let her go.
Once you have found her, never let her go!

Analysis of this text using the methods outlined in Chapter 6 yields the following.[149]

Sentence Structure

The punctuation in the vocal line of the published sheet music (reproduced above) has numerous errors, and moreover does not give a true representation of how the lyrics will be heard as sentences. That would look more like this for the first three stanzas (the last two are more accurate):

Some enchanted evening you may see a stranger.
You may see a stranger across a crowded room.
And somehow you know, you know even then
That somewhere you'll see her again and again.

Some enchanted evening someone may be laughing.
You may hear her laughing across a crowded room.
And night after night, as strange as it seems,
The sound of her laughter will sing in your dreams.

Who can explain it? Who can tell you why?
Fools give you reasons. Wise men never try.

---

[149] The context of the scene, in which Emile and Nellie are reminiscing about their first meeting, should aid in the decoding of the song, although as we shall see it hardly needs the assistance.

The bridge ("Who can explain it?...) consists of simple questions and statements; the coda ("Once you have found her…") is an imperative statement with one subordinate clause. Neither of these should be at all difficult to process.

The sentences in the first two stanzas likewise each have one or two subordinate clauses, but are otherwise straightforward. The fourth stanza (the final A section), however, is one long sentence containing several subordinate clauses. While there is nothing tricky about the sentence, the sheer length of it may add to the difficulty of processing the language, adversely affecting intelligibility.

Surprisal

Only one word in this lyric does not appear in the list of the 5,000 most commonly used English words: *enchanted,* which comes in at 15,555. However, its effective frequency is probably much higher than that, as it is polysyllabic.

Not only does the song eschew rare words, but it is formed largely of extremely common ones. 69% of the words in this lyric are from the 500 most common words in the language; 38% are from the top 100. As more frequent words are likely to be more intelligible, this gives the song a huge intelligibility boost from the starting line.

Unexpected words in their context might include *sing* ("The sound of her laughter will *sing* in your dreams"), *call* ("When you feel her *call* you across a crowded room"), and *dream* ("Or all through your life you may *dream* all alone"). These might cause some difficulties.

Cohort Size

The words that have cohort sizes of twelve or more common words (after the first vowel except as noted) are: *some*, *enchanted*, *see*, *and*, *somewhere*, *someone*, *be*, *laughing*, *hear*, *after*, *as*, *seems*, *of*, *laughter*, *will*, *can*, *explain*, *tell*, *men*, *love*, *feel*, *to*, *own*, *alone* (cohort size 16 after 1st consonant), and *let*. This is about a third of all the words used in the lyric; however, reflecting the preponderance of extremely common words noted above, 18 of these 25 words are the most common words in their cohorts.

Phonetic ambiguity

• There are only three words that start with /b/, /d/ or /g/ in this lyric: *dream* (which appears twice, once as a verb and once in its plural noun form as *dreams*), *give* and *go. Bream* is a sufficiently obscure word that I think we can discount it here; *go* could conceivably be misheard as *doe,* although the context ("*never let her…")* should give *go* a major boost.

• There are no words containing /b/, /d/ or /g/ in a non-initial position that could cause confusion with another word if these phonemes were confused.

• Substitution among /p/, /t/ and /k/: *try* could be misheard as *pry. True* could be misheard as *crew. Call* could be misheard as *pall* or *tall.*

• Words that begin with unvoiced fricatives include *some, somewhere* and *someone*; *strange* and *stranger*; *see, hear, seems, sound, sing, fly, through* and *found.* However, none of these are set particularly high or loudly.

•Words beginning with consonants followed by front vowels include *see, stranger, then, that, her, be, laughing, hear, seems, laughter, dreams, can, tell, reasons, wise, men, never, when, feel, make, dream, let.*

• Words containing a back vowel followed by a consonant include *across, crowded, room, sound, fools, call, all, found.*

<u>"Difficult" words</u>

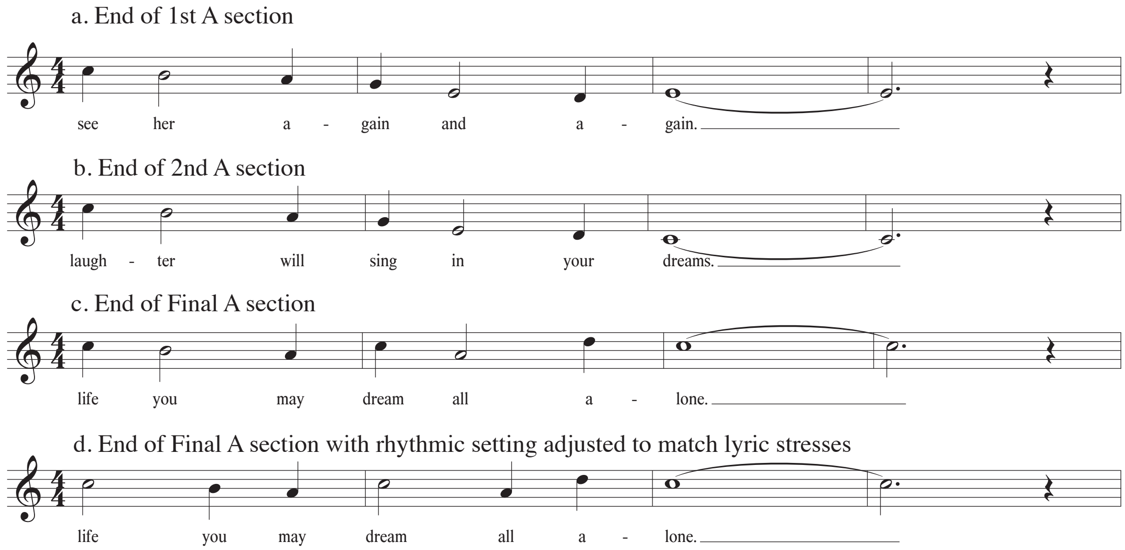
Call appears three times as potentially problematic in this analysis, as does *dream* (twice if we discount any potential confusion with *bream*). Words that appear twice are *enchanted, see, let, be, laughing, hear*, *seems, laughter, tell*, *men*, and *feel.* Words that appear once are *sing, some*, *and*, *somewhere*, *someone*, *after*, *as*, *of*, *will*, *can*, *explain*, *love*, *to*, *own*, *alone*, *go, try, true, stranger, then, that, her, dreams, can, reasons, wise, never, when, across, crowded, room, sound, fools, all, found* and *make.* Any of these words that are given especially problematic settings will be noted in the musical analysis below.

<u>Metric setting</u>

Rodgers has set the three A sections identically, rhythmically speaking. However, in the third A section the metric stresses are different from the other two A sections for one line; hence, there is a mismatch in the stress pattern (see Example 12.1a-d). The words *you* and *all* in Example 12.1c receive stress in their sentences, and would thus be better matched if they were set on beat 3 (see Example 12.1d).[150] There may be a cost to intelligibility in this passage.

[150] I have often heard this adjustment made in concert performances of this song.

Example 12.1. “Some Enchanted Evening” (vocal line only) from *South Pacific* by Richard Rodgers and Oscar Hammerstein II.



Word Groups

The same line in the final A section is the only spot to disrupt word groups; the word groups “you may dream” and “all alone” are distorted somewhat by the metric setting, emphasizing the false word groups “life you” and “may dream all,” at a potential cost to intelligibility

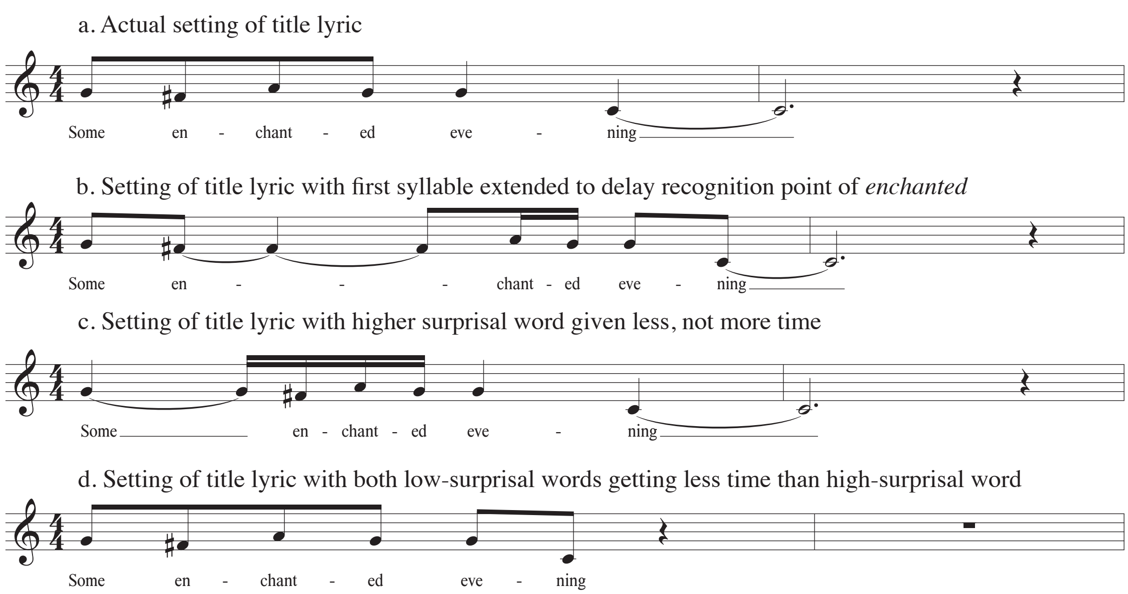
Tessitura

As this excerpt stays safely within the normal baritone range, tessitura is not a factor in intelligibility.

High Surprisal Words

Enchanted is, as noted above, the only word in the song not in the most common 5,000 words in the language. However, the recognition point of the word is at the 2nd vowel (/æ/), relatively early on; if the setting is sufficiently speechlike, there should be little difficulty processing the word. And indeed, the word *enchanted* takes up more time than *some* (although less than *evening*), and the rhythmic setting is such that we reach the recognition point of the word quickly. Example 12.2 contrasts the setting of the title lyric with three imaginary settings, using the same pitch but different rhythms.

Example 12.2. "Some Enchanted Evening," setting of title lyric (vocal line only) from *South Pacific* by Richard Rodgers and Oscar Hammerstein II, with and without rhythmic distortions affecting intelligibility.



12.2a is Rodgers's actual setting. In 12.2b, the first syllable of *enchanted* is extended, delaying the recognition point of the word. In 12.2c, the word *some* is extended and *enchanted* compressed, reversing the expectation that higher-surprisal words get more time. (In both of these, according to my audeation, intelligibility is noticeably downgraded.) In 12.2d, the word *evening* is shortened to make the rhythm of the entire phrase more speechlike, with an interestingly noticeable increase in intelligibility (although a significant cost in lyricism.)

The other potentially high-surprisal words, *sing, call* and *dream,* have nothing noteworthy or difficult about their settings.

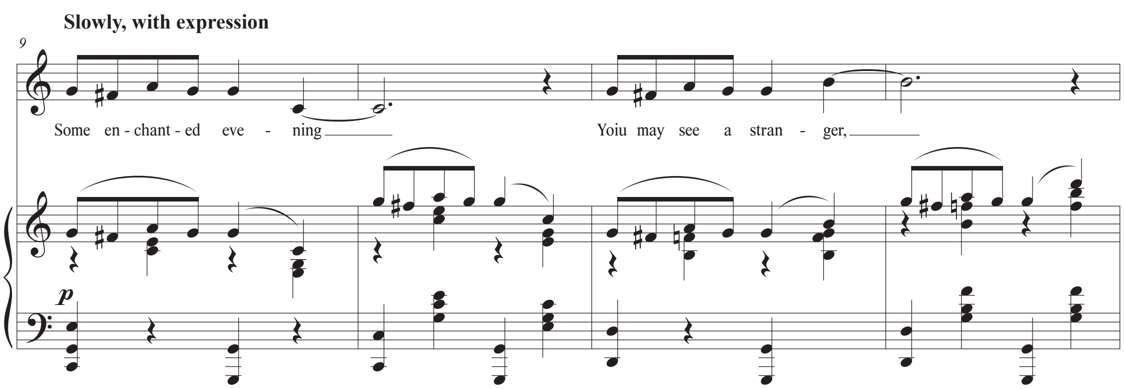
Perception of vowel substitution

The entire piece is written below the pitch where vowel substitution becomes a potential issue.

Chunk Density

The accompaniment consists primarily of bass on 1 and 3 with offbeats on 2 and 4 (which, as a staple of the genre, will come pre-chunked to most listeners), and melodic doubling. There are melodic fills under held notes in the melody; these fall into two categories. One (see Example 12.3) echoes the previous bar's melody, and thus will be perceived as a chunk. The other (see Example 12.4) consists of half- and quarter-note "thumb lines." As the tempo of the song is typically about ♩ = 100, this is not fast enough activity to overload working memory.[151]

Example 12.3. "Some Enchanted Evening," mm9-12 from *South Pacific* by Richard Rodgers and Oscar Hammerstein II.



[151] The fastest tempo I have heard for this piece is actually the Ezio Pinza's rendition on the original Broadway cast recording.

Example 12.4. “Some Enchanted Evening,” mm16-24 from *South Pacific* by Richard Rodgers and Oscar Hammerstein II.

Recognition points

At ♩ = 100, any note held for four beats or more will exceed the two-second limit of working memory. While competition from active accompaniment is not an issue here, the recognition points of the words in question could cause difficulties. The word *try* is held for eight beats, and the word *dreams* and the 2nd syllables of the words *again* and *alone* are held for seven beats. Held for five beats are *know, then, night, seems, side, own* and *go;* held for four beats are the 2nd syllables of *evening, stranger, laughing,* and the word *love.* Let’s consider them in that order. (See Example 12.5.)

Example 12.5. “Some Enchanted Evening,” (vocal line only) from *South Pacific* by Richard Rodgers and Oscar Hammerstein II. Words or syllables that are held longer than two seconds are marked.

The four longest-held words (*try, dreams, again,* and *alone*) will be heavily favored over the rest of the competitors from their cohorts by virtue of their status as the second words of rhyming pairs. Moreover, *dreams* has a nearly nonexistent cohort anyway; *alone* will be recognizable as of the 2nd vowel, the beginning of the held note. *Again* has a couple of potential competitors, but should be assisted by its position as the last word of the idiomatic repetitive phrase "again and again." The recognition point of *try* is not until the end of the word. However, the word *try* is one of only a couple of verbs in its cohort; coming after "wise men never," it would seem to be heavily favored over such relatively high surprisal competitors as *triangulate.*

The five-beat words are either non-rhyming or the first word of a rhyming pair, so they will not receive any recognizability boost from rhyme; however, the songwriters seem to have taken this into account. *Know*'s recognition point is not until the beginning of the next word, but it is immediately repeated in the lyric with a much shorter setting. *Then* has a cohort of only two other words, both of which are much higher surprisal after *even. Night* gets a boost from being part of the idiomatic repetive phrase "night after night" (just as *again* does). *Seems* and *own,* while facing more competion from their cohorts and probably slightly more at risk for unintelligibility, are also the last words of idiomatic phrases: "strange as it seems" and "make her your own."

Of the four-beat words, it is the 2nd syllables of *evening, stranger,* and *laughing* that are held. The recognition points of each of these words is the 2nd vowel; they are each instantly recognizable at the beginning of the held note. While the same is not technically true of *love,* in the context of a 1940s love ballad, *true love* might as well be a two-syllable word; certainly *love* the lowest surprisal possibility after *true* in this context of any word in its cohort. (Compare *true lump* or *true lunch,*for example.)

Despite the presence of numerous words or syllables held out beyond the limit of working memory, there seems to be little risk of a major contribution to unintelligibility from this factor.

Harmonic priming

Here we have an opportunity to interrogate the potential effect of harmonic priming on intelligibility in a context closer to E. Bigand's results (comparing words set on subdominant and tonic chords at cadences). Rodgers has set Hammerstein's repeated word *crowded* on an arrival on IV in each A section; the next arrival, at the authentic cadence at the end of the section, is where the words *again, dreams,* and *alone* fall. While none of these is particularly flagged as difficult, it is suggestive that Rodgers chose the repeated phrase in each A section as the location for a subdominant arrival, as if to position the repetition (increasing intelligibility) to counteract the non-referential harmonic setting (decreasing intelligibility). (See Example 12.6.)

Example 12.6. "Some Enchanted Evening," (vocal line only, 1st A section) from *South Pacific* by Richard Rodgers and Oscar Hammerstein II. Includes Roman numeral analysis.

<u>Conclusions</u>

This song was written for a performer who sang English with a heavy Italian accent. Perhaps a recognition of this obstacle to intelligibility, in addition to that normally imposed by singing, accounts for the care that was taken to construct the majority of this lyric from extremely common words, the most competitive contenders in their cohorts. This would automatically clear away much of the potential difficulty with intelligibility. The most notable aspect of this setting, however, is the way that the potential for trouble with recognition points in the many long notes at the ends of phrases is continually skirted by use of rhyme, repetition, idiomatic phrase, and the sensitive melodic and rhythmic setting of these words. In addition, the potentially intelligibility-reducing effect of the subdominant chord supporting the word *crowded* seems to be balanced by its repetition in each A section.

Despite what seems to be a fairly usual complement of potentially difficult words, the metric mismatch in the last line of the 3rd A section, with its attendant false word groups (*life you* and *may dream all*), is the only misstep in the song concerning intelligibility.

This analysis unveils another facet of the elegance of construction of this evergreen classic.

Chapter 13: Analysis, "Csàrdàs."

"Csàrdàs" (1st stanza, Frischka section) from *Die Fledermaus* by Johann Strauss (music) and Carl Haffner and Richard Genée (libretto, after Henri Meilhac and Ludovic Halévy), English translations by Arthur Anderson (Cranz edition) and Ruth and Thomas Martin (Schirmer edition).

The operetta *Die Fledermaus* ("The Bat") was first heard in Vienna in 1874.[152] It has since become a staple of the light operatic repertory, and is often performed (in the United States) in English translation. Because it exists in numerous English translations, this affords an opportunity to directly compare two of them for their intelligibility. I will be comparing an edition by Cranz from 1911 and the G. Schirmer edition issued in 1986.

The aria "Klänge der Heimat" (familiarly known as "Csàrdàs") was suggested to me as a subject for analysis in the informal poll of singer friends previously mentioned, with the Schirmer edition specifically mentioned. The aria is sung by the soprano Rosalinde in the second act; she is masquerading as a Hungarian countess, and sings a *csàrdàs* (a Hungarian folk dance) to prove her identity, which should help words like *dance* and *whirl* rise to the top of their cohorts.[153] The aria, like the folk dance, begins with a slow section and progresses to a faster section,[154] labeled here "Frischka" or "Friszka." I will be looking at the first stanza of the faster section.

The two texts are as follows[155]:

---

[152] *The New Grove Dictionary of Opera*, "Fledermaus, Die," http://www.oxfordmusiconline.com.proxy.libraries.rutgers.edu/subscriber/article/grove/music/O005972 (accessed August 28, 2017).
[153] Interestingly, the two translations, prepared 75 years apart, seem roughly equivalent in their level of slightly heightened, slightly archaic language.
[154] *Grove Music Online*, "Csárdás," http://www.oxfordmusiconline.com.proxy.libraries.rutgers.edu/subscriber/article/grove/music/06918 (accessed August 28, 2017).
[155] The German original:

Schirmer:	Cranz:
Fiery evening sky, spirits are soaring high. Friends all gather ‘round, hear the Csárdás sound. Lovely gypsy girl, come, dance the merry whirl; Child of Romany, give your heart to me! Fiddles are ringing, hey ya, wildly singing, hey ya! ha!	Joy and a wild unrest fills ev’ry native breast Come join the dance, while you may “Csardas” now they play! Maid with the flashing eyes I claim you as my prize! Come join the dance, while you may Tis a holiday! Ah!

It is worth noting, before proceeding to the rest of the analysis, that the Schirmer translation has 41 words, while the Cranz has 42; but the Cranz omits the last line of the original completely, replacing it simply with a vocalise on *ah.* The Cranz edition uses the same number of words to fill fewer bars of music. Since we know that longer words have a greater effective frequency, the Schirmer edition already has an intelligibility edge.[156]

Sentence Structure

The Schirmer, however, begins with a sentence fragment, or at least a subordinate clause with no verb; what is implied is “[under a] fiery evening sky, spirits are soaring high” or perhaps “[There is a] fiery evening sky. Spirits are soaring high.” In either case, the missing words are likely to derail sentence processing a bit. The next three lines are fairly straightforward imperative sentences; the two direct addresses included (“lovely

---

Feuer, Lebenslust, Schwellt echte Ungarbrust, ,
Heil! Zum Tanze schnell, Csárdas tönt so hell!
Braunes Mägdelein Musst meine Tänz'rin sein
Reich den Arm geschwind, Dunkeläugig Kind!
Zum Fiedelklingen, ho ha, tönt jauchzend Singen: ho ha, ha!

[156] The original German text has only 36 words in these same measures. If the principle that longer words have greater effective frequency works the same way in German, with its abundance of compound words, then the original has a greater chance of being intelligible simply for that reason.

Gypsy girl" and "Child of Romany") should not be difficult to decode as such. The final line is a simple declarative sentence, although its interruption by the exclamation "hey ya" will of course also interfere with the flow of the sentence.

The Cranz, if we fill in and smooth out a bit of wayward punctuation, begins uncomplicatedly enough; the first two sentences (comprising one line each) offer no obstacles. "Csardas now they play," however, is a twisted convolution of a sentence that may confuse a listener and add to the difficulty of language processing. The rest of the stanza mimics the structures of the Schirmer, with a similar direct address ("Maid with the flashing eyes") tied to a declarative sentence rather than an imperative one.

Surprisal

The Schirmer contains six words not in the most common 5,000: *fiery, gypsy, Romany, Csàrdàs, fiddles* and *wildly.* Of these, the two proper nouns, *Romany* and *Csàrdàs,* unsurprisingly also do not appear in the most common 20,000 words. The Cranz contains four not in the top 5,000: *unrest, Csardas, maid,* and *tis. Csardas* is the only one not to appear in the most common 20,000.

In the Schirmer, *whirl*, *ringing* and *singing* are not common continuations of their preceding phrases; however, *whirl* and *singing* should be made much less surprising by their status as second members of rhyming pairs, leaving *ringing* as a potentially difficult word. In the Cranz, *unrest* probably qualifies as high surprisal; the Google Ngram Viewer ranks it as 1,000 less frequent after *wild* than, say *life.* These words may be difficult to decode.

Cohort Size

The Schirmer has twelve words with cohort sizes of twelve common words or more (after the first vowel except as noted): *soaring*, *gather*, *hear*, *lovely*, *gypsy*, *come*, *dance*, *of*, *Romany*, *to*, *fiddles*, and *ringing*. Of these, *to, of*, and *come* are the most common words in their cohorts.

The Cranz edition has only nine: *and*, *unrest*, *fills*, *native*, *come*, *dance*, *with*, *tis,* and *holiday*. *With, come, dance* and *and* are the most common words in their cohorts. *Unrest* in particular has a very large group of competitors, all the words beginning with the prefix *un-*. Even after the 2nd vowel, there are still challengers; the recognition point is not until /s/. Please see the Appendix for discussions of the cohorts and recognition points of each word.

Phonetic ambiguity

• Words that start with /b/, /d/ or /g/: The Schirmer has *gather, girl, dance,* and *give.* The Cranz has *breast, dance.* None of these risks confusion with other words if the phonemes are substituted.

•Substitution among /p/, /t/ and /k/: In the Schirmer, *sky* could be heard as *spy* or *sty*; *heart* could be heard as *harp,* In the Cranz, *play* could be heard as *clay*; *maid* could be heard as *mate*; *prize* could be heard as *cries* or *tries*.

•Words that begin with unvoiced fricatives include (from the Schirmer) *fiery, sky, spirits, soaring, friends, hear, fiddles* and *singing.* From the Cranz: *fills, flashing* and *holiday.* There is no dynamic marking in the vocal line at the beginning of this section; presumably the ***p*** marking from the end of the previous section carries over. Thus, the only potential mishearings are those of these words set above the staff. In the Schirmer, that is *spirits* and *singing;* in the Cranz, *fills.*

• Words beginning with consonants followed by front vowels include *friends, gather, hear, dance, merry,* and *me* from the Schirmer; *native, breast, dance, may, they, play, maid, flashing,* and *claim* from the Cranz.

• Words containing a back vowel followed by a consonant include *soaring* from the Schirmer and *holiday* from the Cranz.

<u>"Difficult" words</u>

From the Schirmer:

Words that appear twice as potentially difficult in this analysis are *hear*, *gather*, *soaring*, *ringing, fiddles*, *gypsy,* and *dance*. Words that appear once are *fiery, Romany, Csàrdàs, wildly, lovely*, *come*, *of*, *Romany*, *to*, *sky, heart, singing, friends, merry,* and *me.*

From the Cranz:

Unrest and *maid* appear three times as potentially difficult in this analysis. Words that appear twice are *fills, native, tis, holiday,* and *play.* Words that appear once are *Csardas, and*, *come*, *dance*, *with*, *prize, breast, dance, may, they, flashing,* and *claim.* The Cranz edition would appear to have an unintelligibility disadvantage. We will keep an eye out for these words' settings in the musical analysis below.

Metric setting

There are no mismatches between the metric stresses and the musical stresses in either translation of this passage. However, each edition alters either the rhythm or the syllable-to-note matching in the service of the translation. Example 13.1 shows the alterations.

Example 13.1. Frischka section of “Csardas” (Klänge der Heimat), 1st stanza, from *Die Fledermaus,* by Johann Strauss (music) and Carl Haffner and Richard Genée (libretto, after Henri Meilhac and Ludovic Halévy), English translations by Arthur Anderson (Cranz edition) and Ruth and Thomas Martin (Schirmer edition). Vocal line only. Boxes indicate alterations in the translated editions, either of the musical rhythm or the the syllable-to-note matching.

The Schirmer smooths out the 32nd notes, used in pairs in the German original as two-note melismas on unstressed syllables, to 16th notes, allowing the lengths of the syllables of the words in those measures to keep something closer to their spoken ratio.

This will certainly increase intelligibility. The Cranz, by contrast, keeps the 32nd notes, but gives them one syllable each. This might work in the first measure, where the 32nd notes carry the short and unstressed words *and a,* but is going to make it very difficult to understand *while you*, as both of these words simply take more time to say than they are being given here. (*With the* falls somewhere in between; it will probably be heard as *with a* anyway.) The Cranz also omits the tie between the 1st and 2nd beats in the 5th measure of this passage, and the parallel measure in the answering phrase, allowing the measure to have three words (*Come join the*) instead of two. Again, as the words more nearly approximate a speech rhythm, this will probably provide a gain in intelligibility.

Word Groups

The integrity of word groups is maintained throughout both settings.

Tessitura

This melody stays primarily within the staff, leaping up to G5 only in a couple of places until the melismatic writing begins at the end of the excerpt. As already noted, *spirits* (Schirmer) and *fills* (Cranz) begin with unvoiced fricatives, and their intelligibility may suffer as a consequence at that altitude; their vowels will also suffer. In the parallel place in the 2nd phrase, the 11th measure of the excerpt, Schirmer has *come* and Cranz has *I,* both of which should fare somewhat better, with their more open vowels. (See Example 13.1, third and eleventh measures.) A less obvious question is whether the words set to two-note melismas from C#5 to G5 (Schirmer: *friends, child*; Cranz: *join*) will be impacted. While we know that consonants and formant transitions suffer above the staff, each of these words contains a nasal or liquid consonant either in the final position or as part of the final consonant cluster (these being the consonants least affected by high tessitura), which ought to mitigate the effect somewhat.

In the final phrase of the excerpt (see Example 13.2), the Schirmer has two words set to melismas circling the top of the staff, *ringing* and *singing.* (The Cranz has this entire phrase simply sung on *ah*.) While the vowels will certain be hard to discern here, and we’ve already noted that the initial consonant of *singing* is vulnerable, it is interesting to note that all of the other consonants are also liquids or nasals: /r/, /ŋ/, the least vulnerable to range-related effects.

Example 13.2. Frischka section of “Csardas” (Klänge der Heimat), mm17-24, from *Die Fledermaus,* by Johann Strauss (music) and Carl Haffner and Richard Genée (libretto, after Henri Meilhac and Ludovic Halévy), English translation by Ruth and Thomas Martin (Schirmer edition). Vocal line only.

<u>**High Surprisal Words**</u>

The high surprisal words in the Schirmer are *fiery, gypsy, Romany, Csàrdàs, fiddles, ringing* and *wildly.* Most of the words in the excerpt are between an eighth note and a quarter note long; of the high surprisal words, *fiery, Romany* and *ringing* are given two full beats; *fiddles* and *wildly* are given one and a half. Only *gypsy and Csàrdàs* are given one beat each. So the translators have managed to give two-thirds of their high surprisal words more time than average, as is helpful with intelligibility.

In the Cranz, the high surprisal words are *unrest, Csardas, maid,* and *tis.* As in the Schirmer, the majority of word durations fall between and eighth and a quarter note; here, *maid* has nearly two full beats; *unrest* has a beat and a half. *Csardas* has one beat; *tis* has a dotted eighth note. Here only half of the high-surprisal words have extra time.

In both translations, the word *Csàrdàs* is the most obscure, and thus most high surprisal word in the text, and its setting will not mitigate that fact. (To be fair to the translators, this is also true in the original German, in which the average word duration is actually longer than it is in either of the English translations.)

Perception of vowel substitution

Schirmer: *Friends* might be heard as *frands*; *merry* as *maary;evening* more like *ivneng*; soaring as *soareng*; *gypsy* as *gepsy*; *fiddles* as *feddles*; *ringing* and *singing* as *rengeng* and *sengeng*.

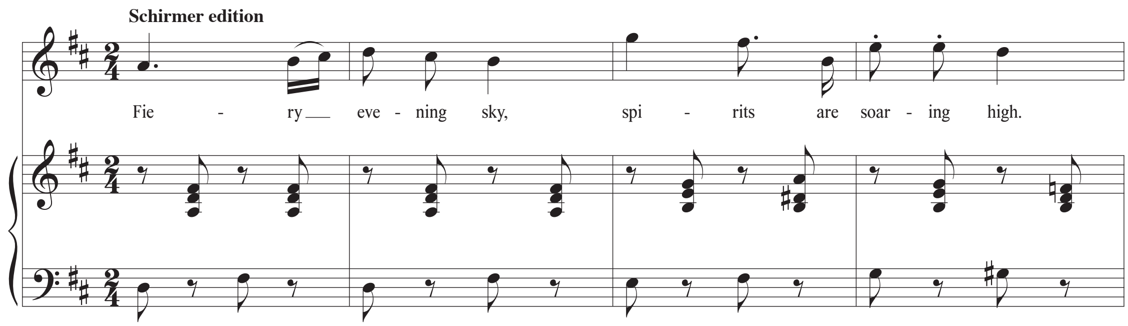
Cranz: *Unrest* might become *unrast*; *every* might be heard as *avry*; *breast* as *brast*; *fills* as *fells*.

This is a potential problem, especially for the Schirmer.

Chunk Density

The *accompaniment* for this section is entirely "oom-pah." (See Example 13.3.) There is no risk of the accompaniment overloading working memory, as this will fade into the the background and not compete with the vocal line. However, see the following paragraph.

Example 13.3. Frischka section of “Csardas” (Klänge der Heimat), mm1-4, from *Die Fledermaus,* by Johann Strauss (music) and Carl Haffner and Richard Genée (libretto, after Henri Meilhac and Ludovic Halévy), English translation by Ruth and Thomas Martin (Schirmer edition).



Recognition points

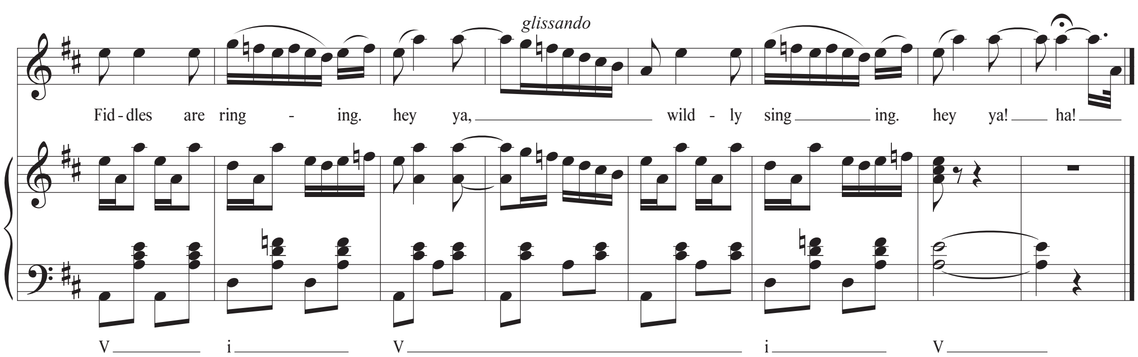
The slowest tempo for this passage I found in a range of recordings was ♩ = 120. At that tempo, the longest-held words in the Schirmer, *ringing* and *singing,* last one second; there is no danger of running into the two-second time limit of working memory. (The non-word *ya*, also in the Schirmer, lasts for 1.5 seconds, and *ha* is set on a fermata.) (See Example 13.2 above.) However, in the vocal line, the note density increases dramatically and suddenly; it is possible that the chunk density of the melisma itself will contribute to the already precarious state of intelligibility for these words.

The entire eight measures from the Schirmer shown in Example 13.2 is simply sung on *ah* in the Cranz. Interestingly, while this the lifts the challenge of intelligibility for the whole phrase, the listener may not know that; since so much of the phrase is set around the pitches above which all vowels start to sound like *ah,* and because there has been little vocalise thus far in the aria, the listener may continue vainly searching for words for at least some of this time.

Harmonic priming

The first two phrases of the stanza are melodically and harmonically identical, and there is nothing in them that could qualify as harmonically unexpected.[157] However, the third phrase, a dominant prolongation, alternates between V and i, the borrowed minor tonic; while not unusual for the exoticist "Hungarian" genre, it probably qualifies as harmonically unexpected in its context. The first appearance of this d minor chord (in the Schirmer edition) is on the word *ringing,* which already has several potential strikes against it. It is repeated for the word *singing* four measures later. (See Example 13.4.)

Example 13.4. Frischka section of "Csardas" (Klänge der Heimat), mm17-24, from *Die Fledermaus,* by Johann Strauss (music) and Carl Haffner and Richard Genée (libretto, after Henri Meilhac and Ludovic Halévy), English translation by Ruth and Thomas Martin (Schirmer edition).



[157] The progression is: I I | ii V4/3/ii | ii6 vii°/V | V7 vii°/vi | vi vii°/V | V7 V7 | I I |

Conclusions

As satisfying as it would be to proclaim one of these translations the clear winner when it comes to intelligibility, the devil's bargain that any opera translator must make in trying to balance sense and singability is obvious in both of them.

Both managed to make some choices that would mitigate intelligibility problems; the words on the two-note melismas, which jump up above the staff, contain liquid consonants, which are less vulnerable to mishearing above the staff than others. In other aspects the two make a sharper comparison. The Schirmer edition has an edge because of its use of more polysyllabic words (and of its decision to alter the 32nd note rhythms to 16th notes). The Cranz edition's decision to make the last line a vocalise seems perhaps to have been a wise one given the likely difficulties with the Schirmer's choice of text on the melismas (*ringing* and *singing* especially), especially as the melismas themselves and their harmonic settings will likely make intelligibility more difficult. The Cranz has more high surprisal words, and not in rhythmic positions which will allow them time to be understood; the Schirmer has more words for which perceived vowel substitution could be problematic.

This analysis shows, however, that the choices translators make can indeed make a difference to intelligibility. While the perfect translation of a song text is probably a mirage, the techniques used in these analyses could help a translator know which of their inevitable tradeoffs will support intelligibility and which will work against it.

Afterword

Jane Ginsborg and Philip Fine identified four categories of factors that influence the intelligibility of sung text: performer-related, listener-related, environment-related and words/music related. A composer usually has direct control only over the fourth category. In this paper I set out to determine what exactly the members of this fourth category were, and what information was available that could help a composer and/or librettist maximize the intelligibility of an English-language text setting for the classically trained voice.

The published works by and for composers on this subject are minimal, not very specific, and only moderately helpful, so I turned first to specialists in the singing voice for explanations of what it is about classically trained singing that makes it so difficult to understand, and to empirical musicologists for recently gleaned information about how this plays out in real world (or at least laboratory test) situations. This in itself is an education for a composer; knowing that formant transitions carry a great deal of the information that makes phonemes identifiable, for instance, or that liquids are the consonants most resistant to range-related mishearing, is extremely useful. So is the idea that the unexpectedness of its harmonic setting can influence a word's intelligibility.

It turns out, however, that there is far more useful information on this subject to be gleaned from disciplines not often studied by composers. Poring through papers from phonetics, acoustics, cognitive science and psycholinguistics yielded a cornucopia of potentially useful insights.

From phonetics, we learned more about the aspects of spoken language that are missing or altered in singing, giving additional insight into the difficulties of decoding sung language. From studies in acoustics we gained some understanding of how some phonemes or combinations of phonemes are more vulnerable than others to being misheard, especially through the sonic interference of piano or orchestral accompaniment, analogous to noise in these researchers' experiments. From cognitive science, we learned about the limits of working memory, both temporal and informational. From psycholinguistics, we learned about the Cohort Model, in which phonemes activate neurological competition among stored words, and about surprisal, in which unexpected words require more neurological processing than expected ones, allowing us to think differently about why some words may simply be harder to understand in certain settings than others.

It is in combining these disciplines' approaches to the issues surrounding intelligibility with the musicologists' insights and composers' intuitions examined in the early chapters that I believe the contributions of this paper lie. This area is ripe for an interdisciplinary methodology; the Wolfson Heuristic for Analysis of Textsetting is, I hope, a worthy first step along that road. The analyses that make up the bulk of the paper have demonstrated how this interdisciplinary approach can aid in the understanding of composer-controlled causes of unintelligibility in classically sung English. We now have a better understanding of exactly why pieces like Adès' "Five Fathoms Deep" and Barber's "Sea-Snatch" are so challenging to make intelligible for even the best performers. We have also gained some insight into some of the ways that composers such as Britten, Sondheim and Rodgers have found to bypass or finesse the difficulties examined here; all three of these composers' piece examined here contain ingenious workarounds of potential pitfalls in intelligibility, identified as such here for the first time.

As interesting as these analyses can be, my hope for this work is that it is useful to the future composer of English language opera and art song. The principles and ideas discussed here could easily become part of compositional pedagogy; they could be condensed into a form no more onerous to absorb than, say, the fingerings for the orchestral strings, and comprise the content of no more than a handful of lessons. The development of the "instinct and experience" that Thomson insists be the guide for text selection could be accelerated; the "sensitivity to the most appropriate syllables for the pitches within a phrase" that Barker recommends composers acquire could be based on research and fact.

Appendix: Cohort Sizes

The following words have cohorts of twelve words in general usage or more:

<u>"Dido's Lament"</u>

Hand has a much more common competitor in *have* and its cognates; however, the syntactic context (after *thy*) makes those words grammatically impossible. The recognition point of the word is the beginning of the final consonant cluster, /n/.

More's cohort includes all of the words beginning with /mar/ (moral, mortal, morsel, etc.). However, the recognition point is the beginning of the next word, *I*, since there are no English words beginning with that combination of phonemes.

But has a large cohort, but it is by far the most common word of the group. The recognition point (differentiating it from competitors such as *butter* or *butler*) is at the first phoneme of the next word.

Death is also the most common word in its cohort. It is recognizable at the final consonant.

Earth has some serious competition from other, at least as common words in its cohort. It is recognizable at the final consonant.

Remember is distinguished from its large cohort at the second vowel.

Forget is distinguished from *forgo* at the second vowel.

<u>"Green Finch and Linnet Bird"</u>

Finch is not a common word, but this is a good example of a word that will be vaulted to the top of its cohort by its context. Its cohort will be narrowed down immensely at the /n/, although its recognition point is the final consonant. It is worth noting here that, at the specified tempo of quarter note equals 112, the word *finch* is almost exactly at a comfortable speaking tempo (whereas other words in the same line, such as *and* and *linnet*, are sustained a little longer than they would be in speech).

And has a huge cohort after the first consonant, but as it is the 3rd most common word in the English language, it outcompetes them all handily. Moreover, it is unstressed in the setting, whereas all the competitors have the first syllable stressed; this would seem to handicap the competition even further.

Bird has a large cohort, including five other words that appear in the most frequently used 5,000. However, it too will be assisted by context.

Can is the most common word in its cohort. However, even at the final consonant a sizable cohort is still in play. Even after the first phoneme of the next word, *canyon* is still a possible competitor, not resolved to *can you* until the vowel of *you* (According to the Google Ngram Viewer, *can you* is three times as common as *canyon*, though.)

Sitting (or at least *sit*) is also the most common word in its cohort, although there are eight others in the top 5,000. However, the /t/ narrows the field down to *sit* and its cognates, and the recognition point is the 2nd vowel.

Wing has a very large cohort, with several more common competitors (*will, wind, with, winter, window*). However, it may also be assisted by the avian context. Furthermore, it is the 2nd member of a rhyming pair.[158]

Outside also has a very large cohort, mostly consisting of all of other words beginning with the prefix "out." It is distinguished from *outsell*, etc. at the 2nd vowel, and from *outsize* at the final consonant.

Beckoning has several more common competitors; however, the cohort is narrowed down to *beck* at the /k/, and the recognition point of the word is the 2nd vowel.

Just is by far the most common word in its cohort. The cohort narrows down to *justice* at /s/, and the recognition point is the beginning of the next word.

Beyond has a huge consort, but its recognition point is the /j/.

Remain is similarly situated; but even at the 2nd vowel, there is still a sizeable cohort (*remade, remail, remake*), and the recognition point is not until the final consonant. It is possible that this word will suffer from its cohort size.

Staring is the second most common of its cohort (only *step* is more common). Its recognition point, however, is the 2nd vowel of the initial diphthong.

At is a very common word, though not as common as its cohort members *am* and *and*; the /t/ shrinks the cohort to a few members, from which it is distinguished at the beginning of the next word.

---

[158] While the effect of rhyme on surprisal has not been studied, common sense would suggest that it would have the effect of drastically narrowing in the cohort of the rhyming word. (Collister and Huron (2008) only studied the unusual case of immediately rhyming words, not the more common case in which a rhyming word comes in a predictable place in the music.)

Rain has a robust cohort of equally common words; it is distinguished from all but *range* and its own cognates at /n/, but the recognition point is the beginning of the next word. This is another one that may be difficult because of cohort size, although it should get a small boost from the internal rhyme with *remain.*

By has a huge cohort because it is homophonic with the prefix *bi-*; its recognition point is the beginning of the next word.

<u>"Sea-Snatch"</u>

Has is a member of a moderately large cohort, although several words in the cohort are cognates (*have, had, hath*). With its cognates, it is the most common word in its cohort. Its recognition point is the final consonant.

Us has a huge cohort, containing (among others) every word beginning with un- or up- as a prefix, but it is the second most common word in its cohort (after *of*). Its recognition point is the final consonant.

O (an alternate spelling of *oh*) has in its cohort every word that begins with /oʊ/, a large gathering. It is the fourth most common word in the cohort, following *only, own* and *open.* Its recognition point is actually midway into the next word, as it is only distinguished from *oak* at the vowel of *king.*

King has a moderately large cohort. It is nearly equivalent in frequency to *kiss*; *kill, kick* and *kitchen* are more frequent. Its recognition point is the final consonant.

Of is a member of the same cohort as *us.* It is one of the five most common words in the English language. The recognition point is the final consonant.

Kingdom is a member of the same cohort as *king.* Its recognition point is the /d/.

Heaven competes with the much more common *head* (and all its combination forms) and *help*, and the equally common *hello*. *Hell* is also more common than *heaven*; make of it what you will. The recognition point is /v/.

Wind is a member of the same consort as *wing* (from "Green Finch and Linnet Bird"). *With* and *window* are its only more common competitors. The consort narrows drastically at /n/, and it is distinguished from *window* at the beginning of the next word.

Consumed has a very large cohort of words that also begin with the prefix *con-*. However, only *concern, connect,* and *conduct* are notably more common. It is distinguished from *consult, consist* and others at the second vowel.

As competes with every word that begins with /æ/. Its primary competitors are the even more common *and, am,* and *at.* Its recognition point is the final consonant.

Timber has a large cohort at the first vowel, a handful of which are more common, but it is narrowed down to *timpani* and *timid* at /m/. It is distinguished from these words at /b/.

Devoured has a huge consort of all words with the prefix *de-*. It is narrowed to a handful at /v/ and distinuished from *devout* at the /r/.

By has a huge cohort because it is homophonic with the prefix *bi-*; its recognition point is the beginning of the next word.

Crimson has a moderately large cohort, but only a handful of them are more common. It is distinguished from all but *crimp* at the /m/, and from *crimp* at the /z/.

<u>"Hymn"</u>

And is by far the most common word in its very large cohort, even if we include the first consonant.

Huntress has a cohort of 16, but only four of those appear in the 5,000 most common words, and only two are more common than the cognate *hunter. Huntress* is distinguished from *hunter* at /r/.

Fair has a moderately large cohort, and competes with several other common words; however, it is closer to the top of the list than most of them. It is distinguished from *fairy/ferry* at the beginning of the next word.

Sun has a huge cohort because of all its compound forms (as well as over 20 other words). It is distinguished from all those compound forms (and "sunk") at the beginning of the next word.

To is another extremely common word with a large cohort. Its closest competitors are its homophones *two* and *too.*

Seated has a huge cohort, mostly of them compounds of *sea.* There are three more common competitors: *see*, *seem*, and *sea*. Recognition point is the 2nd vowel.

Silver has a moderately large cohort, including several more common competitors. It is distinguished from "silvan" at 2nd vowel, from all others at /v/.

State has a cohort of 13. Its recognition point is /t/.

Manner has a large cohort, although it narrows quite a bit at /n/. It is distinguished from *man* at the 2nd vowel, but still competes with its less common homophone *manor.*

Hesperus competes after the 1st vowel with *head* and all its combination forms. Interestingly, though, its recognition point is very nearly the first /s/—the cohort narrows down only to a couple of other proper nouns, such as *Hester.*

Entreats has a gigantic consort, as *end-* is a very common prefix. Even if we look at the consort after *entr-,* it's still pretty sizable, although it does narrow down to a more manageable size (about seven) if we only include members of that group with an unstressed first syllable. The recognition point is /i/.

Goddess has to compete with the myriad combination forms of *god,* as well as a handful of other words; but only two of them (*golf* and *got*) are noticeably more common. Recognition point at 2nd vowel.

Excellently has a cohort of about 1500 words, more than 50 of which are among the 5,000 most common words. Distinguished from all except cognates at /l/.

<u>"Five Fathoms Deep"</u>

Fathom, an unusual word itself, has a cohort of at least 20, of which ten are among the 5,000 most common words. Its recognition point is the 2nd consonant.

Pearls has a very large cohort by virtue of the prefix *per*-, although only a few of these words are in the most common 5,000. Its recognition point is /l/.

His and *him* are both members of moderately large cohort of about 16. There are several other common words in the cohort, as well. Both words have recognition points at the final consonant.

Mortal is distinguished from several more common competitors at /t/; its true recognition point (separating it from *mortar*) is the final consonant.

Same is distinguished from its moderately large cohort at the final consonant.

Coral's recognition point is the 2nd vowel. Its cohort is at least nineteen strong.

Suffered and *something* have a very large cohort partly by virtue of the prefix *sub*, which turns up ten times in the most common 5,000 words, adding to the ten other cohort members already there. The recognition point is the 2nd vowel.

Rich and *ring* have a fairly large cohort, but only a couple of more common competitors. Their recognition points are their final consonants.

And, as noted above, is the most common word in a large cohort. However, since word-stress cues are missing in this setting, it has more competition than it did in "Green Finch and Linnet Bird."

Hourly normally has a small cohort by virtue of its initial glottal stop. However, as this will probably be omitted when sung, it will also be competing with all the combination forms that begin with *out* (there are 11 in the most common 5,000 words.) /l/ is the recognition point.

Can has a very large cohort, and its recognition point is the beginning of the subsequent word.

Ding has a very large cohort, due in part to words with the prefix *dis*-, of which there are over 30 in the most common 5,000 words. Its recognition point is the final consonant.

Some Enchanted Evening

Some (and its compounds *somewhere* and *someone*) has a large cohort, including all words with the prefix *sub*-, but it is by far the most common in the cohort. (See entry for *suffered* above.) It is distinguished from its compounds at the start of the next word. *Somewhere* and *someone* are distinguished from each other at the 2nd vowel.

Enchanted's cohort is huge by virtue of including all words beginning with the prefix *en-,* but is narrowed down to only enchain at [tʃ]; recognition point is at /æ/.

See is likewise the most common word in its very large cohort (see entry for *seated* above). Its recognition point is the beginning of the next word.

A has as its cohort any word that begins with an unstressed vowel—a very large group. Its recognition point is really the point at which there is enough information about the next word to separate the two; in the case of *a stranger*, it is distinguished from *astray* at /n/, and then from *a strain* at /dj/ and *a strange* at the 2nd vowel of *stranger.* In the case of *a crowded,* it is the /aʊ/ that distinguishes it from *across* and all its single-word competitors beginning with unstressed vowels, as none of these have the /aʊ/ diphthong.

And is another word that is the most frequently used member of a large cohort.

Again is a member of the same large cohort as *a.* At the 2nd vowel it is distinguished from all but *against* and the combinations *a guess, a guest,* and *a ghetto.*

Be is the most common word in its very large cohort, made up mostly of words containing the prefix *be-*. *Be laughing* is distinguished from *believe, below* et al. at the /æ/ of *laughing.*

Laughing and its cognate *laughter* have a large cohort, but only *land* and *last* are more common. Recognition point is at /f/.

After is distinguished from all of its large cohort but *aft* at /f/.

As is a member of the same cohort as *after.* Recognition point is the final consonant.

Seems is the 2nd most common member of the large cohort it shares with *see*. Recognition point is the final consonant.

Of is the most common member of its large cohort. Recognition point is the final consonant.

Will is the 2nd most common word in its cohort (after *with*). Recognition point is /l/.

Can is the most common word in its cohort. Even at the final consonant a sizable cohort is still in play. However, as *can* is in an unstressed metric position here, the beginning of the next word, *tell,* is the recognition point; such words as *canter* do not share the metric structure.

Tell's cohort includes all words with the prefix *tele-*. Its recognition point is the beginning of the next word.

Men is the most common word in its large cohort. Its recognition point is the beginning of the next word.

Love has a cohort just over our threshold at size 14. Recognition point is the final consonant. It is worth noting that its position in the music on an unstressed beat (as the 2nd word of the noun phrase *true love*) essentially eliminates the polysyllabic members of the cohort.

Feel is the most common word in its large cohort. Its recognition point is the final consonant.

To is the most common word in its large cohort. The recognition point is the beginning of the next word.

Own is the 2nd most common word in its cohort (after *only*). Its recognition point is the beginning of the next word.

Alone is the 3rd most common word in its large cohort (after *along* and *allow*) Recognition point is the 2nd vowel.

Let is the most common word in its large cohort. Distinguished from all except *lettuce* at /t/.

Klänge der Heimat (Csàrdàs)

Schirmer edition:

Soaring has a moderately large cohort, with a couple of more common competitors (*so* and *sort*). Recognition point, distinguishing it from the smaller cohort starting with /soʊr/ is at the 2nd vowel.

Gather has a very large cohort, but only one more common competitor (*gas*). Recognition point is at the 2nd consonant.

Hear has a moderately large cohort, but is the most common member. Recognition point is the second vowel of the diphthong.

Lovely has a cohort of fourteen, but only two more common competitors: *luck* and *lunch*. The cohort is narrowed down to *love* and its combination forms and cognates at /v/; the recognition point, distinguishing it from *loveless* and *lovelorn,* would be the 2nd vowel.

Gypsy also has a cohort of fourteen, and while it is not common, neither are any of its competitors except *gym*. It is distinguished from all but *gypsum* at /p/, and from that at the 2nd vowel.

Come is by far the most common word in its moderately large cohort. It is distinguished from all but *comfort* and its cognates at /m/, and from those at the beginning of the next word.

Dance's only more common competitors in its cohort of 19 are *dad* and *damage.* Its recognition point is /s/.

Of is the most common word in its huge cohort. The recognition point is /v/.

Romany has a cohort of 19, and several much more common competitors. It is distinguished from all except *Roman* at the 2nd vowel, and from that at the 3rd vowel.

To is the most common word in its cohort of 18. The recognition point is the beginning of the next word.

Fiddles has a cohort of 20, with several much more common competitors. The recognition point is /d/, however, early in the word.

Ringing has a very large cohort, but has only one more common competitor: *risen*. Recognition point is at 2nd vowel for cognates, at /ŋ/ for all others.

Cranz edition:

Unrest is a member of the giant cohort of words with the prefix *un-,* of which there are ten in the 5,000 most common English words (*unrest* itself comes in at 9201). At the 2nd vowel, the cohort narrows down to *unread, unready* and *unrented.* It is distinguished from these at /s/.

Fills is one of the more common words in the cohort it shares with *fiddles*. Its recognition point is the final consonant.

Native has a cohort of 15, with two more common competitors: *name* and *nation.* Its recognition point is the 2^{nd} vowel.

Come is by far the most common word in its moderately large cohort. It is distinguished from all but *comfort* and its cognates at /m/, and from those at the beginning of the next word.

Dance's only more common competitors in its cohort of 19 are *dad* and *damage.* Its recognition point is /s/.

With is the most common word in its very large cohort. Recognition point is the final consonant.

Tis has several more common competitors in its large cohort. Recognition point is the final consonant.

Holiday has only one more common competitor in its cohort of 16: *hot*. Its recognition point is the 2^{nd} vowel.

BIBLIOGRAPHY

Altman, Gerry T. M. *The Ascent of Babel.* New York: Oxford University Press, 1997.

Barker, Paul. *Composing for Voice.* New York: Routledge, 2004.

Benolken, Martha S., and Charles E. Swanson. "The effect of pitch-related changes on the perception of sung vowels." *Journal of the Acoustical Society of America* 87 (1990), 1781-1785.

Berz, William L. "Working Memory in Music: A Theoretical Model." *Music Perception* 12 (1995), 353-364.

Bigand, Emmanuel, Barbara Tillmann, Bénédicte Poulin, Daniel A. D'Adamo, and François Madurell. "The effect of harmonic context on phoneme monitoring in vocal music." *Cognition* 81 (2001), B11-B20.

Burton, Strang, Rose-Marie Déchaine and Eric Vatikiotis-Bateson. *Linguistics for Dummies.* Mississauga: John Wiley and Sons Canada, Ltd, 2012.

Chen, Zhijian and Nelson Cowan. "Chunk Limits and Length Limits in Immediate Recall: A Reconciliation." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31 (2005), 1235–1249

Cohen, Allen, and Steven L. Rosenhaus. *Writing Musical Theater.* New York: Palgrave MacMillan, 2006.

Collister, Lauren B. and David Huron. "Comparison of Word Intelligibility in Spoken and Sung Phrases." *Empirical Musicology Review* 3 (2008), 109-125.

Condit-Schultz, Nathaniel and David Huron. "Catching the Lyrics: Intelligibility in Twelve Song Genres." *Music Perception: An Interdisciplinary Journal* 32 (2015), 470-483.

Cowan, Nelson, and Candice C. Morey, Zhijian Chen, Amanda L. Glichrist and J. Scott Saults. "Theory and Measurement of Working Memory Capacity Limits." *The Psychology of Learning and Motivation* 49 (2008), 49-105.

Demberg, Vera, Asad B. Sayeed, Philip J. Gorinski, and Nikolaos Engonopolous. "Syntactic surprisal affects spoken word duration in conversational contexts." In *Proceedings of Sixth ISCA Workshop on Speech Synthesis*, 294-299. ISCA, 2007.

Di Carlo, Nicole Scotto. "Effect of multifactorial constraints on intelligibility of opera singing." *Journal of Singing* 63 (2007), 443-455, 559-67.

Federmeier, Kara D. "Thinking ahead: The role and roots of prediction in language comprehension." *Psychophysiology* 44 (2007), 491-505.

Fine, Philip A., and Jane Ginsborg. "Making myself understood: perceived factors affecting the intelligibility of sung text." *Frontiers in Psychology* 5 (2014), 1-13.

Fine, Philip, Jane Ginsborg and Chris Barlow. "The influence of listeners' singing experience and the number of singers on the understanding of sung text." *International Symposium on Performance Science* (2009), 51-56.

Grodner, Daniel and Edward Gibson. "Consequences of the Serial Nature of Linguistic Input for Sentenial [sic] Complexity." *Cognitive Science* 29 (2005) 261–290.

Hale, John. "Uncertainty about the rest of the sentence." *Cognitive Science* 30 (2006), 643-672.

Hale, John. *Automaton Theories of Human Sentence Comprehension.* Stanford: CSLI Publications, 2014.

Hale, John T. "Information-theoretical complexity metrics." Unpublished draft of review article, 2015.

Hale, John T., David E. Lutz, Wen-Ming Luh, and Jonathan R. Brennan. "Modeling fMRI time courses with linguistic structure at various grain sizes." *Proceedings of CMCL 2015,* 89-97.

Halle, John. , *Text, Tune and Metrical Form,* http://johnhalle.com/musical.writing.technical/technical.htm (accessed 2/18/16).

Howes, Davis. "On the Relation between the Intelligibility and Frequency of Occurrence of English Words." *The Journal of the Acoustical Society of America* 29 (1957), 296-305.

Howie, John and Pierre Delatorre. "An Experimental Study of the effect Pitch on the Intelligibility of Vowels." *NATS Bulletin* 18 (1962), 6-9.

Jackendoff, Ray. "Parallels and nonparallels between language and music." *Music Perception* 26 (2009), 195-204.

Johnson, Randolph, David Huron and Lauren Collister. "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." *Empirical Musicology Review* 9 (2014), 2-20.

Katz, William F. *Phonetics for Dummies.* Hoboken: John Wiley and Sons, Inc., 2013.

Knudsen, Vern O. "The hearing of speech in auditoriums." *The Journal of the Acoustical Society of America* 1 (1929), 56-82.

Kolinsky, Régine, Pascale Lidji, Isabelle Peretz, Mireille Besson and José Morais. "Processing interactions between phonology and melody: Vowels sing but consonants speak." *Cognition* 112 (2009), 1-20.

Krause, Jean C. and Louis D. Braida. "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility." *The Journal of the Acoustical Society of America* 112 (2002), 2165-2172.

Levy, Roger. "Expectation-based syntactic comprehension." *Cognition* 106 (2008) 1126-1177.

Levy, Roger. "Memory and surprisal in human sentence comprehension." *Sentence processing* (2013), 78-114.

Marslen-Wilson, William D., "Functional parallelism in spoken word recognition." *Cognition* 25 (1987), 71-102.

Miller, George A., and Patricia E. Nicely. "An Analysis of Perceptual Confusions Among Some English Consonants." *The Journal of the Acoustical Society of America* 27 (1955), 338-352.

Nábēlek, Anna K., Tomasz R. Letowski, and Frances M. Tucker. "Reverberant overlap- and self-masking in consonant identification." *The Journal of the Acoustical Society of America* 86 (1989, 1259-1264.

Norman-Haignere, Sam, Nancy G. Kanwisher and Josh H. McDermott. "Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition." *Neuron* 88 (2015, 1281-1296.

Patel, Aniruddh D. "Syntactic processing in language and music: different cognitive operations, similar neural resources?" *Music Perception* 16 (1998), 27-42.

Patel, Aniruddh D. "Language, music, syntax and the brain." *Nature Neuroscience* 6 (2003), 674-681.

Pellegrino, François, J. Farinas and J.-L. Rouas. "Automatic Estimation of Speaking Rate in Multilingual Spontaneous Speech." *ISCA Archive, Speech Prosody 2004, Nara, Japan*. http://ronnetsell.net/html/rate_of_speech_and_comms_effic.html, (accessed 5/20/17).

Pellegrino, François, Christophe Coupé, Egidio Marsico. "Across-Language Perspective on Speech Information Rate." *Language* 87 (September 2011): 539-558.

Poulin-Charronnat, Bénédicte, Emmanuel Bigand, François Madurell, and Ronald Peereman. "Musical structure modulates semantic priming in vocal music." *Cognition* 94 (2005), B67-B78.

Reali, Florencia and Morten H. Christiansen. "Processing of relative clauses is made easier by frequency of occurrence." *Journal of Memory and Language* 57 (2007) 1–23.

Roark, Brian, Asaf Bachrach, Carlos Cardenas, and Christophe Pallier. "Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing." In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing,* 324-333. Association for Computational Linguistics, 2009.

Rosenzweig, Mark R. and Leo Postman. "Intelligibility as a function of frequency of usage." *Journal of Experimental Psychology* 54 (1957), 412-422.

Rubenstein, Herbert, and Irwin Pollack. "Word Predictability and Intelligibility." *Journal of Verbal Learning and Verbal Behavior* 2 (1963), 147-158.

Schön, Daniele, Reyna Gordon, Aurélie Campagne, Cyrille Magne, Corine Astésano, Jean-Luc Anton, Mireille Besson. "Similar cerebral networks in language, music and song perception." *NeuroImage* 51 (2010), 450–461.

Smith, John and Joe Wolfe. "Wagner's music is even better than it sounds: implications of vowel-pitch matching for intelligibility and ease of singing." *The Second International Conference on Music Communication Science* (2009), 92-95.

Strange, Winifred, James L. Jenkins and Thomas L. Johnson. "Dynamic specification of coarticulated vowels." *The Journal of the Acoustical Society of America* 74 (1983), 695-705.

Sundberg, Johan. "Acoustic and Psychoacoustic effects of vocal vibrato." *STL-QPSR* 35 (1994), 45-68.

Sundberg, Johann and Camilla Romedahl. "Text Intelligibility and the Singer's Formant—A Relationship?" *Journal of Voice* 23 (2009), 539-545.

Taft, Marcus, and Gail Hambly. "Exploring the cohort model of spoken word recognition." *Cognition* 22 (1986), 259-282.
Thomson, Virgil. *Music with Words.* New Haven: Yale University Press, 1989.

Titze, Ingo. "Why is the Verbal Message Less Intelligible in Singing than in Speech?" *NATS Bulletin* 38 (1982), 37.

Triplett, William M.. "An Investigation Concerning Vowel Sounds on High Pitches." *NATS Bulletin* 23 (1967), 6, 8, 50.

Wang, Marilyn D. and Robert C. Bilger. "Consonant confusions in noise: a study of perceptual features." *The Journal of the Acoustical Society of America* 54 (1973), 1248-1266.

Background and Analytical Observations on *Heaven's Gate*

Background

Heaven's Gate was written to be the capstone of a cycle of chamber operas on questions of religious faith called *The Faith Operas,* which I had begun several years ago. *Heaven's Gate* joined *Maya's Ark* (2010), for mezzo and baritone; *Rapture* (2012), for soprano and mezzo; and *A Fine Invention* (2014), for soprano and tenor. All are based loosely on real events. *Maya's Ark* tells the story of a woman who built an ark in the parking lot of her church, and imagines the moment when the pastor of that church asks her what she's planning to do with all that scrap lumber. *Rapture* watches a woman expecting to be taken up bodily to heaven in the spring of 2011, while her adult daughter waits to pick up the pieces when it doesn't happen. *A Fine Invention* is about a Christian Scientist couple with a gravely ill child, and what happens to their relationship when one of them suggests taking their daughter to a hospital. Director and dramaturg Experience Bryon, who was an invaluable resource during the processes of writing all of these libretti, characterized them, respectively, as "faith restored," "faith destroyed," and "faith challenged."

Each of these is between 10-15 minutes long. In order to make the cycle a complete evening, my new chamber opera needed to be between 40 and 45 minutes long; in order for it to be an easily producible one, the new piece needed to use the same four singers as the existing pieces. These, then, were the conditions that needed to be met by *Heaven's Gate.*[159]

---

[159] I had planned throughout this process to write a fourth short opera for the two men, so that each singer would have the same number of roles to play over the course of the evening. However, the men came to

Dramaturgy

Heaven's Gate is (again, loosely) based on the story of the 1990s cult of that name, which gained posthumous notoriety after all of its members committed suicide, apparently convinced that they would be reborn into new, alien bodies aboard a supposed spaceship that was following the Hale-Bopp comet (which dominated Earth's skies for a few months in 1997). When I began researching the libretto, I had their story conflated with that of Jonestown, the other famous late 20th century mass cult suicide, and the source of the expression "to drink the koolaid." However, the Heaven's Gate cult's website is still up (www.heavensgate.com), and many references to it can be found online, including some message board posts from former cult members who left before the mass suicide. The opera's details about the cult are drawn from these sources, including their eschatology, their love of *Star Trek,* their web design business, their categorization of Major and Minor Offenses. The division of their suicide into two groups a couple of days apart comes from newspaper reports.

The challenge then became how best to tell this story with only four singers. After deliberation, I assigned the cast as follows: Bo, the cult's leader (baritone); Tom, his young right-hand man (tenor); Tina (mezzo) and Jennifer (soprano), senior and junior members of the cult respectively. The conflicts at work would be between Bo and Tom, over Tom's efficacy as Bo's lieutenant, and between Tom and Jennifer, whose doctrinally mandated denial of her sexuality is not going especially well. Tina was thus going to be

---

dominate *Heaven's Gate* to such an extent that adding more stage time for them would have unbalanced the evening, not balanced it.

less important to the drama; to make up for this, I gave her a critical role in the exposition and a major character aria.

The major dramaturgical hurdle was the depiction of a cult of 39 members with a cast of four. Accordingly, the only scenes in which large numbers of the cult are “present” are the two suicide scenes; in those scenes they are indicated by the crushing, one after the other, of the paper cups that have been poured for them to drink poison from.

The scene breakdown/plot synopsis is as follows (spoiler alert):

SCENE 1: Tom is conversing with someone from outside the cult in an online chat room, proselytizing. We learn that the cult believes that bodies are like vehicles, and can be changed.

SCENE 2: Bo is leading the cult in the equivalent of a church service/recruiting session. We learn more about the cult’s beliefs, including that Bo is not actually human, but an alien visitor from the Next Level Above Human in a human body. The recruiting aspect does not go well; Bo blames Tom.

SCENE 3: Tina tells Bo of a report she’s heard, that there’s a spaceship following the comet. Bo immediately decides this is the sign he’s been waiting for, that it is time for him to escort the rest of the cult back to the Next Level. He announces it to the rest of the cult.

SCENE 4: Tom and Jennifer receive the news, Tom with satisfaction and Jennifer with dismay: she has joined the cult to shut down her libido, and is horrified to find it suddenly resurgent.

SCENE 5: Tom asks Tina for help with a web design client who’s upset at being abandoned. She offers to take over, and sings about her disgust with the world and her joy at soon being able to leave it behind; she ends by sending the client an extremely rude email.

SCENE 6: While working with Tom, Jennifer succumbs to her body’s needs and tries to seduce Tom. Tom rejects her.

SCENE 7: The cult is gathered. Bo leads the first wave of suicides, leaving Tom to lead the second.

SCENE 8: The remainder of the cult is gathered. Tom leads the second wave, including Tina and Jennifer, but at the last minute pours his drink onto the floor.

SCENE 9: Tom has realized that he, too, is a visitor from the Next Level, sent to finish Bo's work.

Musical Unity

As this was a longer piece of music than any of my short operas, and indeed longer than any single work I had previously attempted, I was very attuned to the need to create a musical unity across all nine scenes. Accordingly, in writing the libretto from an outline similar to the one above, I made a point of reusing certain lines of text throughout the opera, to allow them to use similar music or (in some cases) function as recurring motives. Tom quotes several lines of his own aria from Scene 1 at the beginning of Scene 4; the text at the end of Bo's first aria in Scene 2 ("Follow me!") also ends the opera; Tom quotes some of that same Scene 2 aria of Bo's at the beginning of Scene 9.

In addition, the phrase "Next Level," which occurs frequently in the libretto, is always set the same way, melodically, rhythmically, and (usually) harmonically; melodically, this is also (with one exception) true of the phrase "Kingdom of Heaven." Settings of the phrase "the Next Level Above Human" are also similar to each other; they contain the same melodic setting of the phrase "Next Level," and similar rhythms, but may differ slightly thereafter. See Figure 1.

	Sc.1	Sc.2	Sc.3	Sc.4	Sc.5	Sc.6	Sc.7	Sc.8	Sc. 9
Next Level	1	3		2		1			2
Next Level Above Human	1	2							
Kingdom of Heaven	1	6	1	1		1		1	1

Figure 1. Occurrences of the recurring text/melodic phrases “Next Level,” “Next Level Above Human,” and “Kingdom of Heaven” broken down by scene.

There are other, internal musical references that bind the piece together. Here are a few of them:

•The accompaniment motive that begins Jennifer’s aria in Scene 4 and recurs throughout, also eventually appearing in the vocal line, thus has a strong association with Jennifer’s struggle to control her libido. It forms the basis of the music accompanying Jennifer’s attempted seduction of Tom in Scene 6, and the aftermath.

•The motive that opens the opera eventually turns out to be the melody for Bo’s call of “Follow me,” echoed at the end of the evening by Tom and the rest of the cast. This melody also occurs in the instrumental interlude at the beginning of Scene 2.

•The melody that accompanies Tina pouring out the drinks in Scene 7 comes from Bo’s “recruitment” aria in Scene 2.

•The ascending and descending minor third scalar motive introduced in Scene 1 features prominently in the processionals that form the bulk of Scenes 7 and 8.

•The “typing” figure of repeated notes introduced in Scene 1, which can be interpreted as Tom’s chat room conversation partner typing back to him, is used in the instrumental interludes that begin Scenes 6 and 8.

Structure

The entire opera can be characterized as a loosely constructed arch; some aspects of the arch are musical, others are dramaturgical. Scene 1 begins with the “Follow me” motive, and then moves to Tom’s online conversation; Scene 9 begins with an echo of that online conversation, using the same musical material, and ends with “Follow me” (after also referencing Scene 2.) A large proportion of Scene 2 is the first segment of Bo’s “recruitment” aria, which is then repeated almost literally by the rest of the cast; this structure is echoed by the twin processionals in Scenes 7 and 8, which are essentially different arrangements of the same musical material. Scenes 2 and 3 concern the imminent departure of the cult for the Next Level; Scenes 7 and 8 depict it. Scenes 4 and 6 concern Jennifer’s struggle with her libido, and use the associated musical material. Scene 5, in the middle, contains Tina’s aria, the only time we hear her speak out, and the only time we hear the compelling reasons one might have for wanting to leave this world for another one.

Musical language

The harmonic vocabulary in general is primarily tertian, although frequently with added “wrong note” colors. Pitch centers are usually strong, but shift often. Traditional tonal progressions are used sparingly; harmonic progression by smooth voice leading is often a guiding principle. (The two-bar vamp that begins Scene 2 and runs through much of Bo's Scene 2 aria is a good example of this.)

The vocal writing emphasizes singability, with a great deal of stepwise motion and easily singable leaps, and extremes of range in all voices saved for expressive climaxes. The rhythmic vocabulary is conservative. These are stylistic choices which I feel are most effective for musical storytelling; they are equally true of *Heaven's Gate*'s three shorter companions.

There is one quote from an outside source. While *Heaven's Gate* is a serious piece, it does have some comic moments; and given the cult's bizarre fascination with "Star Trek," the opportunity to work in Alexander Courage's instantly recognizable fanfare theme was irresistible. It appears (minus its final note) in the instrumental interlude that begins Scene 3, both in inversion and in its original form; similarly as an instrumental fill under the final note of Bo's aria, which ends Scene 3; and in fragmentary form as the setting for Bo's line "It's time for Star Trek!"

Heaven's Gate Scene 1

David Wolfson


8
The bo-dy is not the per-son. All you see of me is words on a screen, but you have
Vln.
Vc.
Pno.
11
no trou-ble be-liev - ing you are con - vers-ing with a per-son. Why can't you ac - cept that the
Vln.
Vc.
Pno.
15
bo-dy is not the per-son? The bo-dy is just a ve-hi-cle for the
Vln.
Vc.
Pno.


18
spi - rit, a ve - hi - cle that can be changed, like you would
Vln.
Vc.
Pno.
20
rit.
A Tempo
trade in a worn - out car for a shin - y new one.
Vln.
Vc.
Pno.
23
rit.
A Little Faster (♩ = 80)
No, it's true. All of us in Hea - ven's Gate,
Vln.
Vc.
Pno.


26
All of us who know the truth, all of us who fol-low Bo, All of us will be leav-ing our
Vln.
mp
Vc.
Pno.
29
ve-hi-cles soon.
We will be tak-en up, our
Vln.
Vc.
Pno.
32
rit.
spi-rits will be tak-en up to new bo-dies, new bo-dies of the Next Le-vel.
Vln.
Vc.
Pno.


A Tempo I°
Slower—quasi recitative
The Next Le-vel A-bove Hu-man.
Vln.
Vc.
Pno.
They are mem-bers of an a-li-en spe-cies, and they will take us in their star -
ship to the True King-dom of Hea-ven.
poco accel.
poco rit.
A Tempo
If you re-cog-nize the truth of this,


48
then you may be one of us.
Vln.
Vc.
Pno.
p
3
52
Is a-ny-one still there?
Is a-ny-one still there?
Vln.
Vc.
Pno.
attacca

Heaven's Gate Scene 2


15
Vln.
Vc.
Pno.
[Scene 2]
20
Bo
mp
Earth is like a gar - den. Hu-mans are the plants. Mem-bers of the
Vln.
Vc.
Pno.
p
24
King - dom of Hea-ven are the gar-den-ers. Ev-'ry two
Vln.
Vc.
Pno.


29
thou - sand years, a small group of hu - mans, the ones who are rea - dy,
Vln.
Vc.
Pno.
33
is gi - ven the seeds of new souls
"souls"
mf
38
Then a mem - ber of the King - dom of Hea - ven sheds his a - li - en bo - dy and
p
3


41
moves in - to a hu - man ves - sel. He ga - thers those
Vln.
Vc.
Pno.
43
hu - mans who have been cho - sen and gent - ly pre - pares them for the Next Le - vel, the
3
Vln.
Vc.
Pno.
46
Next Le - vel A - bove Hu - man. Two thou - sand years a - go, that hu - man
3
f
mf
Vln.
Vc.
Pno.


49
ves-sel was called Je-sus. Now, at this time,
Vln.
Vc.
Pno.
53
mp
rit.
A Tempo
the hu-man ves-sel is me. I shed my a-li-en bo-dy. I left it in the
p
p
p


58
mf
King-dom of Hea-ven. I came to this hu-man bo-dy to ask you:
Vln.
Vc.
Pno.
62
mp
Are you one of those who has a soul?
p
p
p


65
Could you shed your hu - man bo - dy and leave this world be - hind?
Vln.
Vc.
Pno.
68
mf
Do you want to as - cend to the Next Le - vel?
mp
mf
72
p
Fol - low me, fol - low me, fol - low me!


76
Jennifer
mp
Earth is like a gar - den.
Tina
mp
Earth is like a gar - den.
Tom
mp
Earth is like a gar - den.
Bo
mf
On - ly I can lead
Vln.
Vc.
mf
Pno.
mf


80
The King - dom of Hea - ven are the gar - den - ers.
The King - dom of Hea - ven are the gar - den - ers.
The King - dom of Hea - ven are the gar - den - ers.
you where you want to go. On - ly
Vln.
Vc.
Pno.


84
A small group of us is gi - ven the
A small group of us is gi - ven the
A small group of us is gi - ven the
I can fill the hun - ger in your soul.
Vln.
Vc.
Pno.


88
mf
seeds of new souls.
mf
seeds of new souls.
mf
seeds of new souls.
f
The hun - ger in your soul.
Vln.
Vc.
f
Pno.
f
f


91
Then a mem - ber of the King - dom of Hea - ven
Then a mem - ber of the King - dom of Hea - ven
Then a mem - ber of the King - dom of Hea - ven
On - ly a mem - ber of the King - dom of Hea - ven
Vln.
Vc.
Pno.


93
will ga - ther those of us who have been cho - sen and
will ga - ther those of us who have been cho - sen and
will ga - ther those of us who have been cho - sen and
can show you the
Vln.
Vc.
Pno.


95
gent - ly pre - pare us for the Next Le-vel A-bove Hu-man.
gent - ly pre - pare us for the Next Le-vel
gent - ly pre - pare us for the Next Le-vel A-bove Hu-man.
way to the Next - Le-vel.
Vln.
Vc.
Pno.


98
Are you one of us, who has a soul? Are you rea - dy to
Are you one of us, who has a soul? Are you rea - dy to
Are you one of us, who has a soul? Are you rea - dy to
On - ly I! Fol - low
Vln.
Vc.
Pno.


101
leave this world be - hind?
leave this world be - hind?
leave this world be - hind?
Bo
me!
There is not much time left.
Vln.
Vc.
Pno.
104
The
Our time on Earth is near - ly done.
The co - met is com - ing!
Vln.
Vc.
Pno.


107
mp
The co - - - met!
mp
The co-met is com-ing! The co-met!
co-met is com-ing! The co - met!
Vln.
Vc.
Pno.
pizz.
mp
pizz.
mp
mp
111
Bo
mp
The co-met sig-nals the time when the
Vln.
p
Vc.
p
Pno.
p


114
gar - den of Earth will be plowed un - der. The for - ces of E - vil, the
Vln.
Vc.
Pno.
p
f
117
e - ne - mies of Hea - ven, have pro - grammed hu - ma - ni - ty with greed, with con -
cresc.
arco
121
for - mi - ty, with lust and ha - tred and fear. The
mf


126
Freely
A Tempo
p
pro - gram - ming,
p
pro - gram - ming,
p
pro - gram - ming,
gar - den is most - ly weeds. If you can re - ject this pro - gram - ming,
Pno.
p
130
the de - sires of the flesh, nor - mal - cy,
the de - sires of the flesh, nor - mal - cy,
the de - sires of the flesh, nor - mal - cy,
if you can sub - due the de - sires of the flesh, if you can re - ject the ex - pect - a - tion of nor - mal - cy,
Pno.


134
ci - vi - li - za - tion,
ci - vi - li - za - tion,
ci - vi - li - za - tion,
if you can se - pa - rate from so - called ci - vi - li - za - tion, then may - be
"civilization"
Vln.
"civilization"
(vln)
Vc.
Pno.
138
mf
you are one of those who has a soul!
Vln.
mp
Vc.
mp
Pno.
mp


140
May - be you can shed your hu - man bo - dy and
Vln.
Vc.
Pno.
142
leave this world be - hind! May - be
Vln.
Vc.
Pno.


144
to the Next Le - vel!
to the Next Le - vel!
to the Next Le - vel!
you can as - cend to the Next Le - vel!
Vln.
Vc.
Pno.

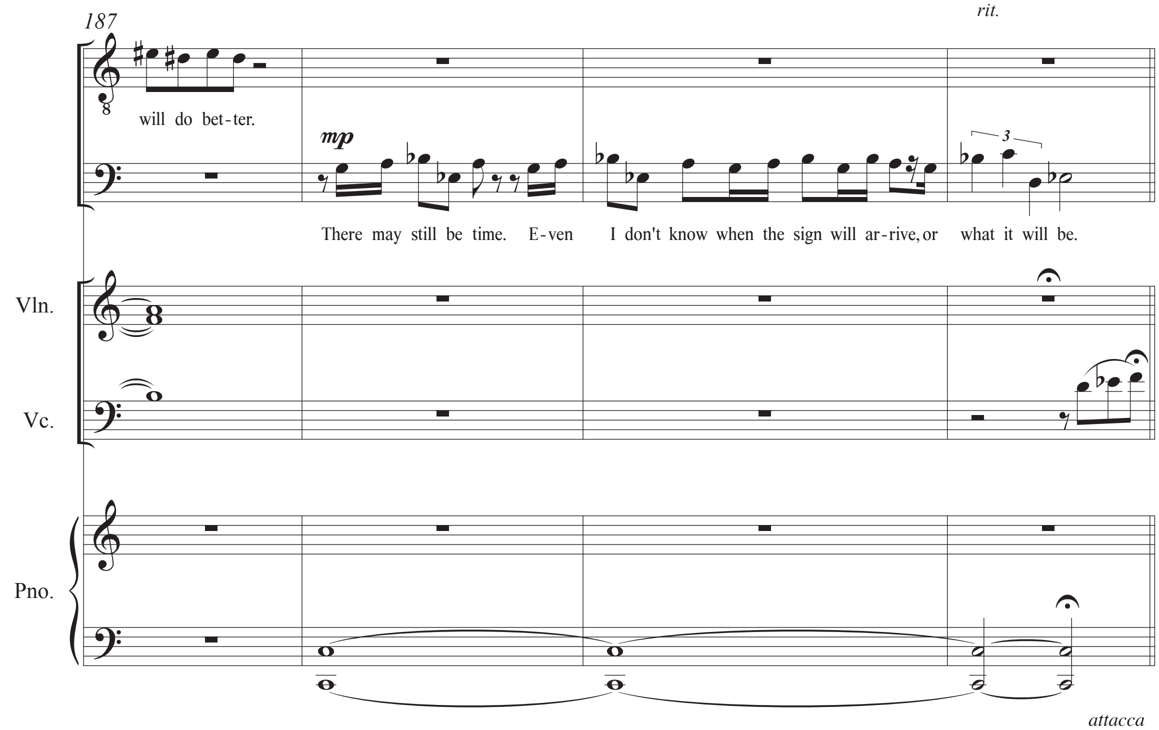

147
mp
There is not much time left! Fol - low
Vln.
mf
mp
p
Vc.
p
Pno.
mf
153
rit.
A Tempo
me, fol - low me, fol - low me!
Vln.
Vc.
Pno.


157
mp
Jennifer
If a-ny-one has a-ny ques - tions...?
Tina
mp
If a-ny-one
Tom
mp
If a-ny-one has a-ny
Pno.
161
rit.
has a-ny ques-tions...?
ques - tions...?
Vln.
(vln)
Vc.
Pno.


Slightly Slower (♩ = 90)
167
Bo
mf
There should have been more peo - ple, Tom. Why so few?
Vln.
mp
Vc.
mp
Pno.
170
Tom
mf
Sev-'ral of our web - site de - sign cli - ents had dead - lines last week.
Why so few?
Vln.
Vc.
Pno.


172
There was on-ly so much time. You your-self said, "Un - til we can leave this cess-pool be-hind, we
we
Vln.
Vc.
Pno.
176
Freely
still must pay the bills."
still must pay the bills." Yes, I said that. But you need to be-come more a-ware of pri-o-ri-ties, Tom.
Vln.
Vc.
Pno.


179
There are so few of you. In all my time in this bo-dy, on-ly thir-ty-nine hu-man souls
Vln.
Vc.
Pno.
183
I'm sor-ry. I
to take back to the King-dom of Hea-ven. You have to do bet-ter.
Vln.
Vc.
Pno.


187
rit.
will do bet-ter.
mp
There may still be time. E-ven I don't know when the sign will ar-rive, or what it will be.
Vln.
Vc.
Pno.
attacca

Heaven's Gate Scene 3


9
I think it is, yes.
por - tant? It's time for Star Trek.
Vln.
Vc.
p
Freely
14
I heard a re-port to-day... ...that an as - tro-no-mer in Tex - as has seen an ob - ject
Yes?


16
A Tempo
mf
fol-low-ing the co-met.
He said a "Sa-turn-like ob-ject."
Fol-low-ing the co-met! What sort of ob-ject? A
Vln.
p
Vc.
p
mp
20
Freely
Sa-turn-like ob-ject?
And a re-mote view-ing ex-pert from a u-ni-ver-si-ty in Vir-gi-nia says it's a


A Tempo
23
star - ship.
A star - ship!
Fol-low-ing the co-met. Could it be...?
Vln.
Vc.
28 rit.
Martial (♩ = 104) (l'istesso)
Vln.
Vc.


32
Bo
mf
This is the sign we've been wait - ing for. This is the mo - ment of truth.
Vln.
mp
Vc.
mp
mp
36
It is time to leave these ve - hi - cles be - hind.
Vln.
Vc.


39
The star - ship fol - low - ing the co - met has been sent for me and
Vln.
Vc.
43
3
you. We have on - ly a mat - ter of days to keep the ren - dez - vous.
Vln.
Vc.


47
mp
We've made our pre - pa - ra - tions. Now, be - fore they pass us
Vln.
Vc.
51
f
mf
3
by, to reach the King - dom of Hea - ven we must let these bo - dies
Vln.
mf
mp
Vc.
mf
mp
mf


56
f
die. We must let these bo - dies
Vln.
Vc.
mf
mf
mf
59
ff
die!
Vln.
Vc.
f
f
3
3
3
3

Heaven's Gate Scene 4


Our spi - rits...
spi - rits will be tak - en up to new bo - dies of the Next Le - vel.
Vc.
And they will take us in their star - ship to the True King - dom of Hea - ven.
Vln.
And they will take us...
And they will take us...
rit.
Vln.
pizz.
Vc.


A Tempo (♩ = 112)
Fi - nal - ly I can leave this bo - dy be -
Vln.
Vc.
arco
ord.
hind, this bo - dy and its want - ing, want - ing
mp
p
mf
mp


mf
28
29
30
31
f
want - ing to wal - low in filth and muck. The feel - ing of the touch - ing,
Vln.
mp
Vc.
mp
mf
mp
mf


32
f
33
f
34
p
35
36
touch - ing, touch - ing, the ca - ress, and the stroke and the
Vln.
mf
p
Vc.
mf
p
3
mf
mf


Faster (♩ = 148)
suck.
Aaaaah!
Not
Vln.
Vc.
now!
Not
still!
Not af - ter all these years of
Vln.
Vc.

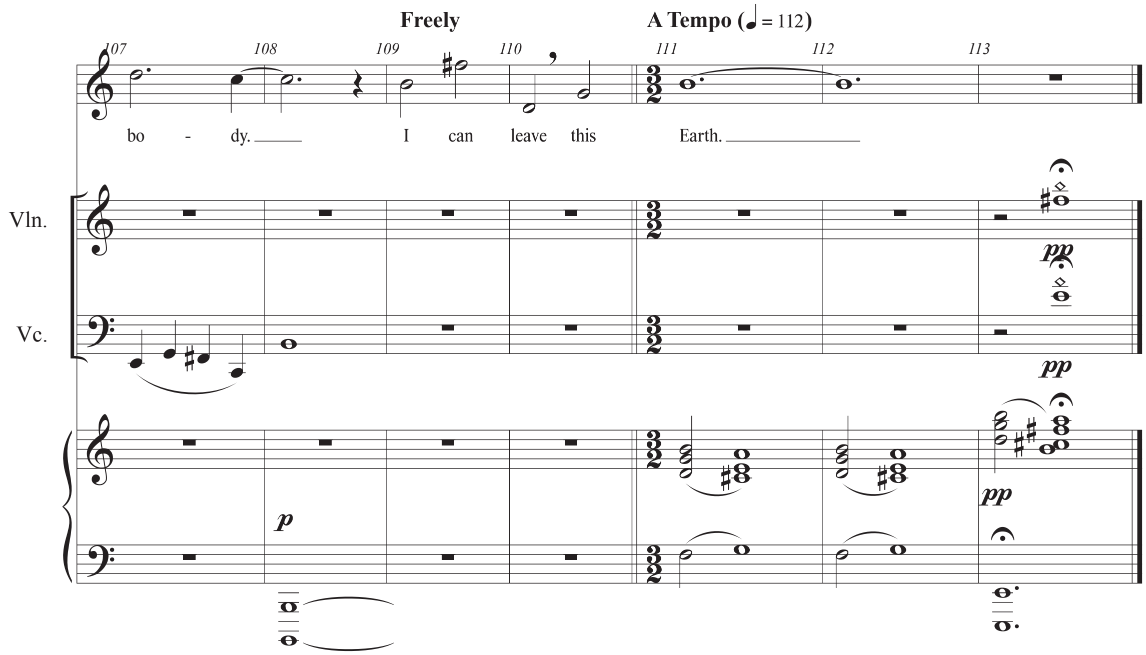

shut-ting it a-way! Of sup-press-ing de-sire
Vln.
Vc.
ev-'ry sin-gle day! Not when it is fi-nal-ly time to
Vc.
go to the Next Le-vel!
Vln.
Vc.


Think a-bout what you know, Jen-ny!
Think a-bout who you are!
All those slut-ty, shame-ful ways are in your past.
All of the rut-ting like a-ni-mals in heat,
treat-ing your-self like a
Vln.
Vc.
mp
mf
p
pp

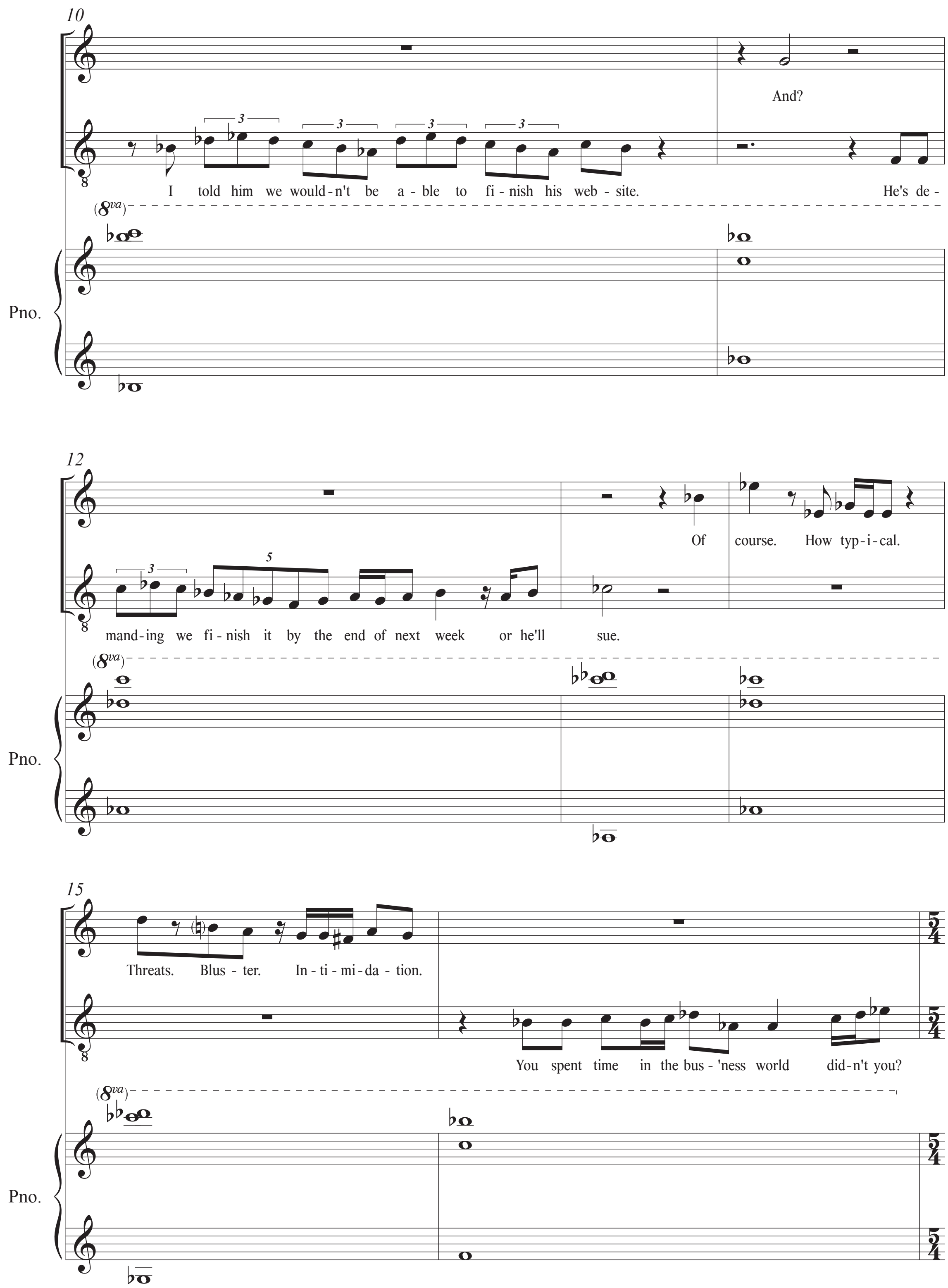

69
70
71
mf
cresc.
piece of meat.
Plea-sure is dis-gust - ing,
plea-sure is dis-gust - ing,
Vln.
p
cresc.
Vc.
p
cresc.
p
cresc.
72
73
74
75
ff
plea-sure is...
Aaah!
Vln.
f
Vc.
f
f


Aaah!
Ah!
Ah!
Vln.
Vc.
Slower (♩ = 112)
p
p


I will be strong. This is not who I am.
Vc.
I can leave this body and its want-ing, want - ing, want - ing...
Vln.
Vc.
Freely
A Tempo (♩ = 112)
I can prove my worth.
I can leave this
Vln.
Vc.


Freely
A Tempo (♩ = 112)
107
108
109
110
111
112
113
bo - dy. I can leave this Earth.
Vln.
Vc.
p
pp

Heaven's Gate Scene 5


10
And?
I told him we would-n't be a-ble to fi-nish his web-site.
He's de-
(8va)
Pno.
12
Of course. How typ-i-cal.
mand-ing we fi-nish it by the end of next week or he'll sue.
(8va)
Pno.
15
Threats. Blus-ter. In-ti-mi-da-tion.
You spent time in the bus-'ness world did-n't you?
(8va)
Pno.


A Tempo (♩ = 104)
17
heavily
Oh yes, I did. Be - fore I met Bo, be - fore I heard the call, I spent eight - y hours a
pizz.
Vln.
Vc.
pizz.
Pno.
21
week or more try - ing to crack the glass ceil - ing.
arco
Lost a hus - band,
Vln.
Vc.
Pno.


25
lost my health, near-ly lost my mind.
Vln.
Vc.
arco
Pno.
29
Tell you what, Tom— let me do it. I
So how do I handle-le this?
Vln.
Vc.
Pno.


34
poco rit.
know you have a lot on your plate.
[he exits]
That's great, Ti - na, thanks.
Vln.
Pno.
39
Seething (♩ = 100)
mf
Tina
I can - not wait to be leav - ing this mo - ney grub - bing
pizz.
Vln.
mp
Vc.
mp
Pno.
mp


43
pit, and these ti - ny - mind - ed peo - ple who are so full of shit,
Vln.
Vc.
Pno.
46
and their pow - er games and their monk - ey ways, Just a few more
Vln.
Vc.
Pno.


50
accel.
Grimly (𝅗𝅥 = 72)
days...
Good
rid-dance to stif-ling
arco
Vln.
Vc.
Pno.
53
so-cial struc-tures.
Good
rid-dance to cor-p'rate rule.
Good
rid-dance to po-li -
Vln.
Vc.
Pno.


58
ti-cians who are ev-'ry kind of fool! Good rid-dance to war and
Vln.
Vc.
Pno.
63
hun-ger. Good rid-dance to hate and fear. Good rid-dance to those who
Vln.
Vc.
Pno.


68
see all this and think "There's pro - fit here!" Good rid - dance to all the
Vln.
Vc.
Pno.
73
mil - lions who sit there all day long, who ne - ver o - pen their
Pno.
77
eyes to see that ev - 'ry - thing is wrong! Ev - 'ry - thing is wrong!
Vln.
Vc.
Pno.


A little slower—in 4
83
mp
On - ly here, on - ly in this place, does ev - 'ry - one see the world for the dis -
Vln.
p
Vc.
Pno.
p
86
grace it is. On - ly here, here at Hea - ven's Gate, is there
Vln.
Vc.
Pno.


89
a - ny - one with the sense to ap - pre - ci - ate me and what I
Vln.
Vc.
Pno.
92
rit.
know. Just a few more days, and then...Good
Vln.
Vc.
Pno.


Grimly (𝅗𝅥 = 72)
96
rid-dance to so-cial in-jus-tice. Good rid-dance to vi-o-lent crime. Good
Vln.
Vc.
Pno.
mf
p
mf
100
rid-dance to a world that cuts off peo-ple in their prime!
Vln.
Vc.
Pno.


104
Good rid-dance to shal-low en-ter-tain-ment. Good rid-dance to fast
Vln.
Vc.
Pno.
108
food! Good rid-dance to all those peo-ple who are ev-'ry kind of rude!
Vln.
Vc.
Pno.


113
rit.
Poco Maestoso
ff
And good rid-dance to sleaz-y cli-ents we don't need a-ny-more.
Vln.
Vc.
f
Pno.
118
mp
We had to pay our bills. That's what you were for.
p
123
accel.
A Tempo (as before)
mp
That is no long-er true, so
mf


127
here's what I have to say to you:
Vc.
Pno.
[she reads back what she has typed]
131
[she types]
Dear Mis-ter Ti-mo-thy:
arco
pizz.
Vln.
Vc.
Pno.
136
[she hits SEND]
FUCK OFF.
Vln.
Vc.
Pno.

Heaven's Gate Scene 6


10
[they laugh]
Are you sure the co-lor's right, though?
I think it's your an-gle.
Vln.
p
13
Look at the mo-ni-tor from down here.
Vln.
mp
16
A little faster (♩ = 72)
mf
You're right. It's fine. Bet-ter than fine. It's great.
Vln.
Vc.
mp


19
mf
Do you think the site will at-tract much traf-fic when we've gone?
In the first few days, may-be.
Vln.
Vc.
mp
22
[TOM rolls his shoulders]
But of course by then it-'ll be too late for all of them.
← 𝅗𝅥 = 𝅘𝅥. → (𝅘𝅥 = 63)
[JENNIFER rubs TOM'S shoulders.]
intense
mp


Jen-ni-fer— what are you do-ing? Sen-su - a - li-ty is one of the ma-jor of-
Vln.
Vc.
Don't be sil-ly. You've been at this for ho-urs. This is-n't sex, it's... phy-si-cal the-
fens-es.
Vln.
Vc.
Freely
- ra-py. Just be-cause we're leav-ing these bo-dies be-hind soon should-n't mean that we ig-nore their


A Tempo
A little faster (♩ = 72)
34
pain.
[TOM acquiesces. Gradually she includes his chest and arms, and it becomes apparent that she is doing this for herself more than for him.]
Vln.
Vc.
mp
38
42
accel.
Tom
mf
Jen-ni-fer, stop.

[...and then suddenly swivels the task chair around, straddles TOM and kisses him, trying to take off his shirt (or hers?). He tries to push her away. She clings to him, and finally he gives a huge shove.]


[She falls to the floor.
He stands up.]
THIS BAR CUT
A little slower (𝅗𝅥 = 72)
55
f
What are you do - ing? What are you do - ing? We leave all
Vln.
Vc.
tr
f
sempre f
mf
61
accel.
things be - hind! Things of the world! Things of the bo - dy!
A little more urgently (𝅗𝅥 = 80)
65
mf
Sex is what mam - mals do, cats or cows or chim-pan-zees, not mem-bers of the
mp


71
mp
King - dom of Hea - ven! We are ce - li - bate, in thought and in deed.
Vln.
p
p
76
f
We are not pri - so - ners of the bo - dy's need. You
Vln.
81
A little slower (𝅗𝅥 = 72)
mp
know this! You know this! You are weak.
Vln.
mf
Vc.
mf
p
mf
p


87
f
I'm not!
You have no self-con-trol. You are just like the rest of them.
Vln.
p
Vc.
91
mp
f
I am!
You are not rea-dy for the Next Le-vel.
Vln.
p


96
Freely
p
I will! I'm sor - ry! Please—
3
You have to do bet - ter.
Vln.
Vc.
p
mf
mf
101
[TOM nods, then exits.]
Slowly (𝅗𝅥 = 60)
don't tell Bo?
Vln.
Vc.
p
p
attacca

Heaven's Gate Scene 7


Solemnly (♩ = 80)
10
Bo
mp
The first group leaves to-day.
p
13
We have been wait-ing for this mo-ment a long time. There are
16
cots sit-ting in the back room on which we can leave our emp-ty ve-hi-cles.
Vc.
p


19
The drink works quick - ly. Once you have downed your cup, go lie down.
Vln.
Vc.
22
You will a - wak - en in your new bo - dy in the star - ship
Vln.
Vc.


A Tempo
poco rit.
25
fol-low-ing the com - et.
I am go-ing with the first
Vln.
Vc.
28
group to ar-range your wel-come ce-le-bra-tion. Tom, I lay up-on you the task of
Vln.
Vc.


32
3
seeing the second group safely to our destination.
Vln.
Vc.
Slightly Faster (♩ = 96)
35
Tom
mp
Is there any final test of readiness? What if one of us has recently com-
p
p


40
mf
mit-ted one of the Ma - jor - Of-fen - ses? What if one of us
Vln.
mp
Vc.
46
f
mp
has not left the world en - ti - re - ly be-hind?
Vln.
mf
mp
Vc.


52
Bo
Freely
mp
That is their wor-ry, Tom, not yours. Let us be-gin.
Vln.
Vc.
p
pp
57
♩ = 104
p
[The procession begins. Every two bars, a "cult member" steps forward to take a drink from BO (he crumples the cup and drops it to the floor).]
63
rit.
Solemnly (♩ = 90)
3
mf
67


71
Vln.
Vc.
mf
mp
f
p
76
81


85
Vln.
Vc.
p
mp
cresc.
89
mf
f
93
ff
p


97
Vln.
Vc.
p
mp
102
rit.
Freely
Bo
In - to your hands, Fa - ther, I com - mend our
[BO drinks]
A Tempo
Slowly, accel.
107
spi - rits.
cresc.


111
Vln.
Vc.
mf
f
ff

Heaven's Gate Scene 8


Solemnly (♩ = 90)
11
Safe tra-vels, and see you in the King-dom of Hea-ven.
Vln.
Pno.
p
[The procession begins again. Every two bars, a "cult member" steps forward
to take a drink from TOM (he crumples the cup and drops it to the floor).]
15
Vln.
Vc.
Pno.
19
Vln.
Vc.
Pno.
mp


24
Vln.
Vc.
Pno.
mp
mp
mp
mp
mf
mf
mf
29
poco rit.
A Tempo
mp
mp
mp
mf
33
p
p


37
Vln.
Vc.
Pno.
[Jennifer grabs cup from Tom and drinks]
[Jennifer exits]
41
[Tom takes cup]
47
cresc.


52
Vln.
Vc.
Pno.
mf
f
ff
[Tom pours drink
on floor]
[scene change]
A Little Slower (♩ = 80)
57
mp
62
attacca

Heaven's Gate Scene 9


10
for - ces of e - vil were too strong for Bo.
Pno.
mp
14
He was un - a - ble to com - lete his
Pno.
17
task. On - ly thir - ty nine souls!
Pno.
l'istesso (♩ = 72)
21
f
I have been sent
Vln.
mf
Vc.
Pno.
mf
mf


from the Next - Le - vel to suc -
Vln.
Vc.
Pno.
24
ceed where Bo failed, to ex - tend hu -
Vln.
Vc.
Pno.


rit.
ma - ni - ty's last chance. You have one more chance!
Vln.
Vc.
Pno.
As in Scene 2 (♩ = 96)
29
p
32
mp
On - ly I can lead you where


35
you want to go. On - ly I can
Vln.
Vc.
p
p
Pno.
39
fill the hun - ger in your soul.
Vln.
Vc.
Pno.

43
mf
On - ly a mem - ber of the King - dom of Hea - ven can show you the way to the
Vln.
mp
p cresc.
Vc.
mp
p cresc.
Pno.
mp
p cresc.
46
ff
Next - Le - vel. On - ly I!
(ah)
ff
pp
ff
pp
ff


51
Jennifer
Fol - low me!
Tina
Fol - low me!
Fol - low me!
Bo
Fol - low me!
Vln.
Vc.
Pno.
55
rit. poco a poco
molto rit.
Fol - low me!
Fol - low me!
Fol - low me!
Fol - low me!
Vln.
Vc.
Pno.


Maestoso (♩ = 80)
59
Fol - low me!
Fol - low me!
Fol - low me!
Fol - low me!
Vln.
Vc.
Pno.