Music-Language Correlations and the "Scotch Snap"

Nicholas Temperley
University of Illinois at Urbana-Champaign

David Temperley
Eastman School of Music, University of Rochester

In this study we examine a rhythmic pattern known as the Scotch Snap (SS): a sixteenth-note on the beat followed by a dotted eighth-note. A musical corpus analysis shows that the SS is common in both Scottish and English songs, but virtually nonexistent in German and Italian songs. We explore possible linguistic correlates for this phenomenon. Our reasoning is that languages in which stressed syllables are often short might tend to favor the SS pattern. The traditional distinction between long and short vowels correlates partly with the SS pattern across languages, but not completely. (German allows short stressed vowels, but the SS pattern is not common in German music.) We then examine the duration of stressed syllables in four modern speech corpora: one British English, one German, and two Italian. British English shows a much higher proportion of very short stressed syllables (less than 100 ms) than the other two languages. Four vowels account for a large proportion of very short stressed syllables in British English, and also constitute a large proportion of SS tokens in our English musical corpus. This is the first study known to us that establishes a correlation between speech rhythms in languages and musical rhythms in the songs of those languages.

Received August 12, 2010, accepted March 1, 2011.

Key words: rhythm, meter, stress, vocal music, music-language connections

It has sometimes been claimed that there are correlations between language and musical style across cultures or nations—that is, that the music of a nation tends to resemble its language in some way. In the words of the musicologist Gerald Abraham, "The nature of a people's language inevitably affects the nature of its music not only in obvious and superficial ways but fundamentally" (1974, p. 62). It is in the domain of rhythm that these effects have most often been observed. Abraham makes a number of specific claims regarding the influence of linguistic rhythm on music. He suggests that the high frequency of "feminine" (strong-weak) phrase endings in Italian melody is due to the prevalence of this pattern in Italian speech. (This is certainly exemplified in traditional Italian recitative, where every line ends with that pattern.) Abraham also proposes that the rarity of upbeats in Hungarian and Czech melodies is due to the prevalence of initial word stress in those languages. As another example, the phonetician Roe-Merrill Heffner claims that "languages with strong stress are likely to have [musical] rhythms of no subtlety whatever; languages which make less use of stress contrast have rhythms which are less obvious. The rhythms of English and French, for example, are in many ways comparable to the rhythms of Sousa and Debussy" (1950, p. 227).

Interesting though they are, Abraham's and Heffner's observations are of an informal, speculative nature; no quantitative evidence is offered, either for the musical phenomena at issue or for their correlation with linguistic phenomena. In general, quantitative support for music-language correlations is difficult to find. In the proceedings of the conference "Music, Language, Speech and Brain" (Sundberg, Nord, & Carlson, 1991), for example, many general analogies between language and music are explored, often supported by quantitative data, but none of them directly compare the music of different cultures. For instance Fant, Kruckenberg, and Nord (1991) compare stress patterns in the reading of Swedish, French, and English prose and poetry, but their analogies with musical rhythm deal only with Swedish verse.

In recent years, rhythmic correspondences between music and language have been subject to more rigorous examination. Particularly notable in this regard is the work of Patel and Daniele (2003). These authors employ a measure of variability in speech rhythm known as the *normalized Pairwise Variability Index*, or nPVI. Given a series of elements (such as syllables or notes), the nPVI measures the degree of variation in duration from one element to the

Music Perception volume 29, issue 1, pp. 51–63. issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2011.29.1.51

next. The nPVI can vary from 0 to 200: a low value (close to 0) means that each element has virtually the same duration as the previous one; a high value (close to 200) means that elements alternate between very short and very long durations. Earlier studies found that languages differ significantly in their nPVI (Grabe & Low, 2002; Ramus, 2002); for example, the nPVI of Dutch and English is significantly higher than that of French and Spanish. (Like most studies of nPVI in language, these studies measured the length of vowels rather than syllables.) Patel and Daniele sought to determine whether these national differences carry over to music. Examining instrumental melodies by British and French composers, and computing the nPVI for each melody using the notated durations, they found a significant difference in nPVI between French and English melodies, in the same direction as the corresponding languages: English showed a higher nPVI (more variable durations) than French (see Figure 1). A subsequent study by Huron and Ollen (2003) replicated these findings with a much larger sample of French and English instrumental themes, as well as many themes by composers of other nationalities. In general, as noted by Patel (2008), the differences between national musical styles found by Huron and Ollen align with the traditional classification of languages as "stress-timed" (with regularly spaced stresses and irregular syllable length) or "syllable-timed" (with regular syllable length): nations with stress-timed languages tend to have higher musical nPVI's.

The findings of these authors offer intriguing evidence for correspondences between musical and linguistic rhythm. One aspect of this work that deserves comment

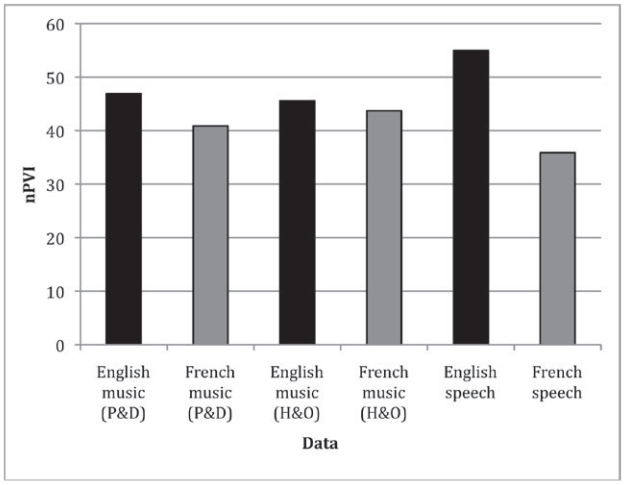


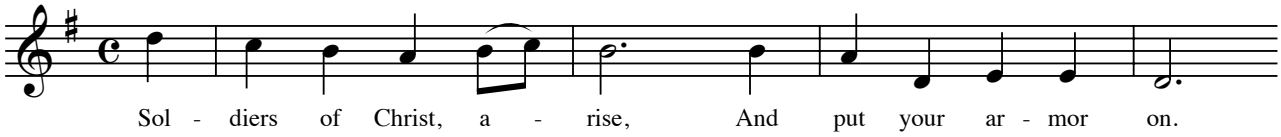
FIGURE 1. nPVI data for English and French music and speech. English music data are from Patel and Daniele (2003) (P&D) and Huron and Ollen (2003) (H&O). Speech data are from Patel, Iversen, and Rosenberg (2006).

is its exclusive focus on instrumental melodies (that is, its exclusion of vocal melodies). Patel and Daniele comment that, "in music based on words...it would be no surprise if musical rhythm reflected linguistic rhythm"; this kind of influence of linguistic rhythm on musical rhythm, they suggest, is "obvious" (2003, p. B42; see also Patel, 2008, pp. 159-160, 165). It is no doubt true that the composition of vocal melodies is often influenced by a desire to match the natural rhythm of the text being set—we might call this the "text-setting factor." By excluding vocal music, Patel and Daniele presumably hoped to eliminate the text-setting factor and thus capture more general influences of linguistic rhythm on a composer's music. While we sympathize with this goal, we believe it is ultimately futile to try to exclude the text-setting factor. There is no doubt that European instrumental music has been greatly influenced by vocal music—partly in cases where instrumental pieces were based on vocal melodies, but also in a more subtle ways, by virtue of the fact that most instrumental composers have heard a great deal of vocal music (and, in many cases, have composed it as well). If linguistic rhythms have influenced vocal music through text-setting, this influence has no doubt been passed on, to some extent, to instrumental music as well.

Thus it seems very likely that correlations between speech and instrumental music are due in part to the mediating effect of vocal music and the text-setting factor. This in no way diminishes the interest of such correlations. Contrary to Patel and Daniele (2003), we believe it is far from obvious that rhythm in vocal music would reflect strong cross-cultural correlations with linguistic rhythm. Rhythm in vocal music is in no way *determined* by the rhythm of the text being set; this is shown by the fact that different settings of the same text can have very different rhythms. Figure 2 shows the beginnings of two hymn tunes written for the same English text. While the two tunes are similar in style—simple, duple-meter melodies designed for congregational singing—their rhythms are completely different. Monk's tune follows the iambic verse meter, while Naylor's comes closer to the rhythms of speech, and at the same time suggests the marching implied by the words.

The relationship between linguistic rhythm and that of vocal music was explored in a recent study by VanHandel and Song (2010). The authors compared 19th-century songs in a stress-timed language (German) and a syllable-timed one (French), and found no difference in rhythmic variability between songs in the two languages. This shows that rhythmic correlations between language and vocal music cannot be taken for granted. (German instrumental music has also been shown to have relatively low rhythmic variability; see

(a) W. H. Monk, "St. Ethelwald" (1861)



(b) E. W. Naylor, "From strength to strength" (1902)

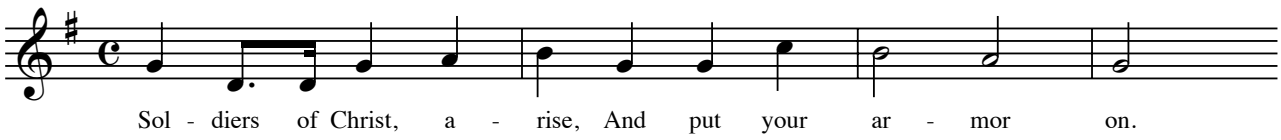


FIGURE 2. Differing settings of the same words.

Patel, 2008, for discussion.) Indeed, in searching for correlations between linguistic and musical rhythm, it makes sense to consider vocal music first, since correlations with language are probably more likely to be found there; if they are found, this justifies the search for them in instrumental music as well.

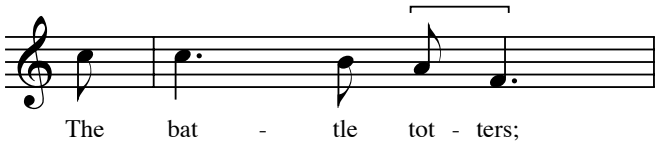
The approach of Patel and Daniele (2003) was to begin with a well-established linguistic difference and search for it in music. In this study, we take the opposite tack: we examine a striking difference in rhythm between British music and that of other European nations, and ask whether any linguistic correlate can be found for it. The difference concerns the so-called "Scotch Snap," a rhythmic pattern involving a short note on a strong beat followed by a longer note on a weak one. Our investigations show that this pattern is widely used in English-language songs, but is virtually nonexistent in Italian and German songs. We then advance two possible linguistic correlates for this phenomenon. First, we consider the conventional distinction between long and short vowels (some languages have this distinction but others do not); this correlates only partly with the musical difference, distinguishing English from Italian, but not from German. We then examine the actual durations of syllables in corpora of English, German, and Italian speech. We find that, while the mean duration of stressed syllables is very similar across the three languages, English has a significantly greater proportion of very short stressed syllables than either Italian or German. We argue that this linguistic difference may offer a plausible explanation for the Scotch snap phenomenon.

THE SCOTCH SNAP

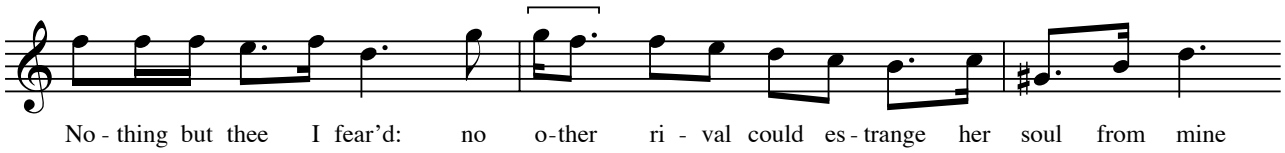
The Scotch Snap (hereafter SS) is a rhythmic figure found in vocal music, typically consisting of a division of a quarter-note tactus in the ratio of 1:3, that is, a sixteenth and a dotted eighth, with one syllable on each note, and with the sixteenth on the beat (see Figure 3 for some examples). It exists in musical notation in many songs of both English and Scottish origin from about 1675 onwards. A less extreme form (an eighth and a dotted quarter, shown in Figure 3A) can be found in England some sixty years earlier, and was in effect similar to the later examples, since the absolute value of durational note symbols became longer during the 17th century (Houle, 1987, p. 8; N. Temperley, 1981). Very possibly the SS was used by singers still earlier but was not notated. It has been associated specifically with Scottish music only since the middle of the eighteenth century (Fiske, 1983; Johnson, 2001; Latham, 2002). No SS can be found in 46 songs collected in *Music of Scotland 1500–1700* (Elliott & Shire, 1975). In more recent times the SS is often found in English-language songs of all kinds (art music as well as folk songs; American as well as English and Scottish ones). (The presence of the pattern in *modern* popular music is a difficult issue, which we will address in the final section of the article.)

Informal observation suggests that the SS pattern is extremely rare in the music of other West European nations. A complicating factor here is the Lombardic Rhythm (LR), which was popular with Italian composers and their imitators in the late 17th and much of the 18th century. The LR also consists of a sixteenth and a dotted eighth, but as a melisma on a single syllable, or in instrumental music, where it is typically slurred (see Figure 4). It can be thought of as a grace or ornament. The composer Johann Joachim Quantz called it the *lombardischen Geschmact* [sic] (Lombardic taste), illustrated it with the slurred groups shown in Figure 5, and said that the manner began in Italy about 1722, but he added that "it seems to resemble Scottish music" (1752, pp. 309–10). It is possible that the LR merged with the SS for a time: Purcell (1659–1695) used both freely. The two are treated as equivalent in many modern sources (Johnson, 2001; Latham, 2002). The

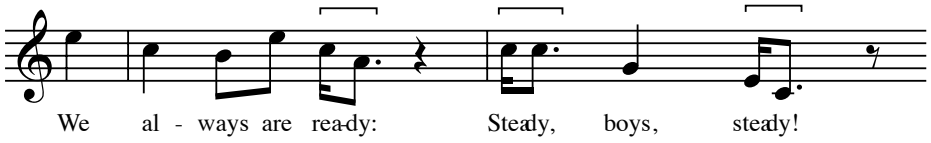
(a) Robert Johnson, from *The Mad Lover*, c.1616



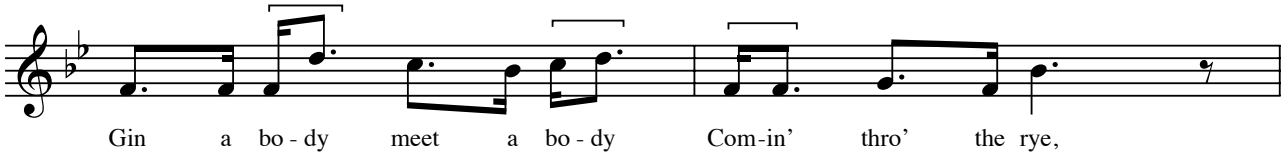
(b) Matthew Locke, from "In a soft vision," 1679



(c) William Boyce, "Heart of Oak," c.1760



(d) Robert Burns, poem set to a Scottish traditional tune, c.1790



(e) George F. Root, "Tramp, tramp, tramp," 1864

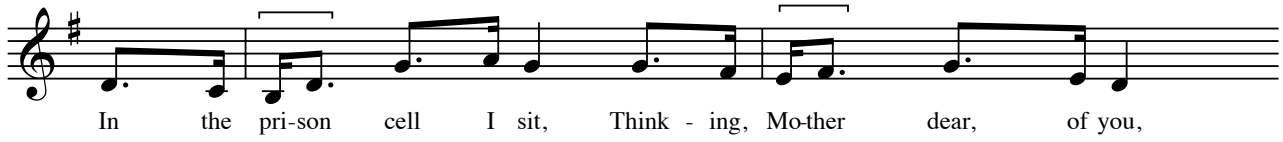


FIGURE 3. Examples of the Scotch Snap are indicated by brackets above the staff.

musicologist Peter Holman (2001) even calls the LR, as found in 18th-century English operas, one of the "elements of British popular song.". But the SS, by our definition, is a purely vocal figure set to two adjacent syllables. It both preceded and long outlived the Europe-wide fashion for the LR.[1]

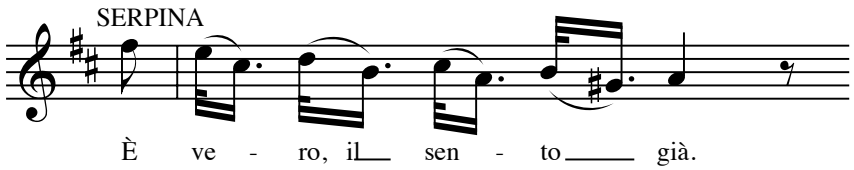
It appears, then, that the SS is much more frequent in English-language music than in music of other European nations. (As we will show, it is more correct to associate the SS with British music generally, rather than with Scottish music in particular; but since the name "Scotch Snap" is widely used, we retain it here.) Our main aim in this paper is to examine whether this difference in musical style can be attributed to differences between the corresponding languages. First, however, we provide some quantitative data regarding the frequency of the SS in musical styles of different nations.

[1] The SS pattern is sometimes imitated in instrumental music, for instance in the strathspey, a Scottish dance that often uses songs played on fiddles to accompany the dance. In instrumental contexts, it is sometimes difficult to distinguish the SS from the LR. In the corpus analysis below, we focus on vocal music, where the SS and the LR can be clearly distinguished.

COUNTING SCOTCH SNAPS IN MUSICAL CORPORA

Since the SS has been found frequently in British songs of the 17th, 18th, and 19th centuries, and in both art songs and folksongs, we chose two popular miscellaneous

(a) Giovanni Battista Pergolesi, *La serva padrona* (1736), No. 7



(b) Mozart, String Quartet in D Minor (K. 471), Trio

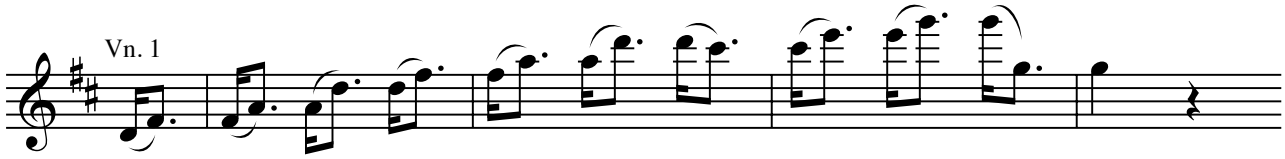


FIGURE 4. Examples of the Lombardic Rhythm.

Quantz (1752), p. 59

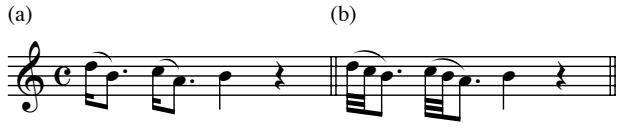


FIGURE 5. Quantz's first two examples of the Lombardic Rhythm.

collections for voice and piano published in the 1870s to represent English and Scottish songs (see Table 1). We were unable to find German or Italian collections of a similarly miscellaneous kind, so we chose in each case two collections, one of art songs and one of folk or popular songs. Only songs in 2/4, 3/4, or 4/4 meter were examined. In compound meters, such as 6/8, tactus divisions in the ratio 1:2 are common (an eighth followed by a quarter), but this pattern appears (from informal observation) to be quite frequent in all European national styles. For simplicity, we excluded compound-meter songs altogether. We examined 100 songs in each language. Where the chosen collections contained more than 100 eligible songs, only the first 100 were counted.

Only the vocal line of each song was examined. Introductory recitatives, da-capos, and repeated sections (whether written out or indicated by signs) were passed over. In strophic songs, only the first stanza was considered, unless rhythmically differing notation was provided for the other stanza(s). In reporting the total number of measures in a song, only those containing music for the voice were counted. We counted every quarter-note tactus that was divided into two notes, each sung to a separate syllable: one a dotted eighth (or an eighth followed by a sixteenth rest), the other a sixteenth. There were thus two types: the regular dotted pair (RDP) in the ratio 3:1, with the dotted eighth coming first and on the beat

TABLE 1. Collections Used for the Scotch Snap Count.

Nation	Collections
England	Hatton, J. L., & Faning, E. (Eds.) (c 1873). *The Songs of England*. 3 volumes. London: Boosey & Co. (First 100 eligible songs counted)
Scotland	Pittman, J., & Brown, C. (Eds.) (c 1878). *The Songs of Scotland*. London: Boosey & Co. (First 100 eligible songs counted)
Germany	Anonymous (1927). *Die Lieder des Zupfgeigenhause Volksweisen*. Leipzig: Hofmeister. (All 50 eligible songs counted) Schubert, F. (1961). *100 Songs*. Sergius Kagan (Ed.). New York: International Music. (First 50 eligible songs counted)
Italy	Anonymous. (1999). *Favorite Neapolitan Songs*. Rome: Edizioni Bideri. (All 27 eligible songs counted) Parisotti, A. (Ed.). (1894). *Anthology of Italian Song of the 17th and 18th Centuries*. 2 volumes. New York: G. Schirmer Inc. (First 73 eligible songs counted)

(as "Gin a" in Figure 3D); and the SS, in the ratio 1:3, with the sixteenth coming first and on the beat (as "body" in Figure 3D). Triplet rhythms, where the tactus is divided in the ratio 2:1 or 1:2, were not counted, nor were slower pairs consisting of a dotted quarter plus an eighth (as in Figure 3A). The results are shown in Table 2.

It is possible that one nationality favors faster rhythms more than another (for example, more sixteenth notes). If that was the case, that nationality might reflect a higher frequency of SS's simply for that reason. It may also be that one nation favors more uneven subdivisions of the tactus (i.e., dotted rhythms in general). It seemed important to eliminate these confounds; it is for this reason

that we counted RDP's as well as SS's. The ratio of SS's to RDP's offers an indication of the prevalence of the SS pattern in particular, in relation to the overall frequency of dotted rhythms. This ratio is also shown in Table 2.

Table 2 shows that the SS occurs about 17 times per 1,000 measures in English songs, or about once every 59 measures. It is significantly more common in Scottish songs. It is nonexistent in the Italian and German corpora; not a single instance of it was found. The significant presence of the SS in English and Scottish songs, and its total absence in German and Italian ones, clearly support our general hypothesis that it is associated with the English language—especially, but not exclusively, with Scottish songs. RDP's are present, and fairly common, in all four of the corpora. (The variation in RDP frequency between the corpora is an interesting issue, but we will not pursue it here.) In all four nationalities, RDP's are much more frequent than SS's. This is not surprising, given the general tendency in Western music for longer notes to be aligned with strong beats (Lerdahl & Jackendoff, 1983; D. Temperley, 2010).

Comparing English with Scottish songs, it is evident that dotted rhythms of both types are much more numerous in the Scottish songs than the English ones. Within the category of dotted pairs, SS's actually form a higher portion in the English than the Scottish, but in absolute terms there are many more SS's per 1,000 measures in the Scottish songs, and there are nearly twice as many Scottish songs as English that contain SS's at all. We conclude that the SS is characteristic of musical settings of the English language in general, but is more marked in Scottish songs. It is also possible, however, that after a connection between the SS and Scottish music had been proposed by Quantz (1752) and others, Scottish composers and editors of Scottish folksong felt encouraged or pressured to use it more often. Since we were unable to find any suitable corpora of Scottish speech, we could not include Scotland in the linguistic comparisons that follow.

TABLE 2. Count of Regular Dotted Pairs (RDP's) and Scotch Snaps (SS's) in Song Collections.

Nation	English	Scottish	German	Italian
No. of songs in corpus	100	100	100	100
No. that contains SS's	18	32	0	0
Total no. of measures counted	4723	1631	4639	3569
No. of RDP's	488	951	720	201
No. of SS's	80	107	0	0
RDP's per 1,000 measures	103.3	583.1	155.2	56.3
SS's per 1,000 measures	16.9	65.6	0	0
SS's per RDP	0.16	0.11	0	0

THE LONG/SHORT VOWEL DISTINCTION

Having confirmed that the SS pattern is far more common in English songs than in German or Italian ones, we sought a linguistic explanation for this phenomenon. The distinctive feature of the SS is a note on a strong beat that is short compared to surrounding notes (or at least, compared to the following note). It is well known that, in Western music at least, there is a strong tendency to align stressed syllables with relatively strong beats of the meter (Baroni, 1991; Halle & Lerdahl, 1993; Palmer & Kelly, 1992). Thus, we reasoned that if a significant proportion of stressed syllables in a language tend (in natural speech) to be short, a natural musical setting of such syllables might result in short notes on strong beats, thus giving rise to the SS pattern.

Do European languages differ with regard to the duration of stressed syllables? One possible answer to this question lies in the distinction between "long" and "short" vowels. In modern phonology, the long/short distinction is a "distinctive feature" that characterizes vowel contrasts within the sound system of a language (Harris, 1994; Lass, 1984). "Short" vowels are not necessarily always shorter than "long" ones in literal duration, and short vowels may differ from their long counterparts in respects other than length (e.g., formant frequency). Nevertheless, the "short" vowels in English do tend to be shorter in duration than the "long" vowels.[2] The six so-called short vowels in British English are those in *cat, pet, pit, pot, cut,* and *put*.[3] Inspecting occurrences of the SS in our English song corpus, we found that it was most commonly used with disyllables such as *pity*, *never*, *mother*, and *bonnie*, where the first syllable is a stressed short vowel. In contrast, Italian does not distinguish phonologically between "short" and "long" vowels (Rebora, 1958). Thus, all stressed syllables in Italian lend themselves readily to long notes or extended melismas, as do stressed "long" vowels in English words (*father*, *easy*).

The fact that English allows short stressed syllables, while Italian does not, offers an explanation for the higher frequency of the SS in English songs. It encounters

[2] Not all phonologists agree that the "short/long" feature is crucial for phonemic contrasts in English; some have argued that the "tense/lax" dimension (with which the short/long dimension is correlated) is more important (Giegerich, 1992; Ladefoged & Johnson, 2011). But even these linguists concede that English vowels differ systematically in length (Ladefoged & Johnson, 2011, p. 99).

[3] The term "British English," as it is commonly used, refers to the dialect (or dialects) of English spoken in England only, not Scotland.

problems in the case of German, however. German, like English, allows stressed short vowels: examples include the first syllables of the words *Bitte, Mutter*, and *treffen* (Hall, 1992; Messinger, 1973). Yet our corpus analysis suggests that the SS pattern is virtually nonexistent in German melodies, as it is in Italian ones. Thus, the short/long vowel distinction, in itself, does not seem to offer a complete explanation for the presence of the SS in some national styles but not others.

COMPARING DURATIONS OF STRESSED SYLLABLES ACROSS LANGUAGES

Since the theoretical long/short vowel distinction did not provide a convincing explanation for the observed cross-cultural differences in SS frequency, we explored another possible explanation. As noted earlier, the long/short distinction is not always reflected in actual syllable durations. Even if both German and English allow categorically "short" stressed vowels, they may differ quite substantially in terms of the actual distribution of syllable durations. We therefore sought to examine the durations of stressed syllables in German, English, and Italian by analyzing corpora of speech in the three languages.

The identification of syllable durations in speech is a complex process. It requires the creation of a phonological representation from the speech signal, the identification of syllable boundaries in this representation, and the identification of precise timepoints for the syllable boundaries. The second of these three steps—the identification of syllable boundaries—is in some cases quite controversial: that is, linguists do not always agree on whether a particular consonant belongs with the previous syllable or the following one (Ladefoged & Johnson, 2011). Also, it is difficult to find corpora that contain multiple languages (or multiple corpora that are precisely comparable). Still, suitable corpora were found for all three languages that seemed comparable enough for present purposes.

For English, we looked for British English rather than American English speech data, as the songs in our corpus were British. (Below we will return to the distinction between British and American English.) We used the Aix-Marsel corpus, taken from BBC radio recordings (Auran, Bouzon, & Hirst, 2004). It is divided into sections representing different kinds of speech (news, commentary, fiction, and so on). One section contains dialogue (e.g., talk shows and interviews) and that section was used here, as it seemed to offer the closest compatibility with the German corpus described below. The corpus includes a phonological transcription, annotated with syllable boundaries (with their timepoints), lexical stress (i.e., distinctions of stress within a word), and other information. The stress markings identify syllables as having either primary stress, secondary stress, or no stress; both primary and secondary stress were taken to indicate a stressed syllable. By this criterion, the corpus contained 4,720 stressed syllables.

For German, we used data from the Verbmobil corpus, a large corpus of spontaneous German speech taken from conversations about appointment scheduling, travel planning, and computer repair (Burger, Weilhammer, Schiel, & Tillmann, 2000). The corpus is phonologically annotated, showing syllable boundaries, their timepoints, and lexical stress information; again, both primary and secondary stresses are shown. The available data contains over one million syllables; we used approximately the first 10% of the data, containing 28,748 stressed syllables.

It appears that there are no large, publicly available corpora of naturally occurring Italian speech with syllable times and stress information. We did, however, find two corpora that, taken together, offer an adequate representation of Italian speech rhythm. The first corpus is a subset of the Italian SPEECON corpus (Siemund, Hoege, Kunzmann, & Marasek, 2000), annotated with syllable duations and stress markings (Abete, Cutugno, Ludusan, & Origlia, 2010). The corpus, which features speakers from various regions of Italy, contains 1,789 stressed syllable tokens. The limitation of the corpus is that the spoken text consists entirely of numbers, from zero to 999,999. The corpus includes 35 different stressed syllables, including instances of all seven of the Italian vowels: [a], [e], [ɛ], [i], [o], [ɔ], and [u].

A second Italian data set was taken from a study by den Os (1988), comparing Italian and Dutch. Three Italian speakers read two short newspaper texts, yielding a total of 511 syllables, including 114 stressed syllables. The texts were labeled with stresses by a phonetician and a native speaker; the native speaker's judgments were used, though a high level of agreement was found between the two. Syllable boundaries and timepoints were identified. This raw data is not provided in the study and is no longer available (den Os, personal communication, March 7, 2010). However, the study does provide some aggregate data: den Os grouped the syllables by the number of phonemes they contained (one through four), and provided the mean and standard deviation of duration for each of the four categories; the data for stressed syllables is shown in Table 3. Thus, comparing our two Italian corpora, the SPEECON corpus features complete data on syllable durations but a limited vocabulary; the den Os corpus features a broader vocabulary but only aggregate duration data. Taken together, the two corpora provide a reasonably satisfactory picture of stressed syllable durations in Italian.

Given these data showing the durations of stressed syllables in English, German, and Italian, our aim was to examine whether English showed a higher frequency

TABLE 3. Duration of Italian Stressed Syllables (from den Os, 1988).

Number of phonemes	Number of tokens (% of total)	Mean duration (ms)	*SD* (ms)
1	0	–	–
2	51	211	46
3	42	252	46
4	7	330	51

of very short syllables. The overall means of duration for stressed syllables in the three languages were very similar: 233 ms for English, 241 ms for German, 246 ms for the Italian SPEECON corpus, and 237 ms for the Italian den Os corpus. (Den Os, 1988, does not report mean syllable duration, but it can be computed, given the means and counts for the four phoneme categories, in Table 3.) It is also of interest to consider the standard deviation for each language. For English, it is 117 ms; for German, it is 114 ms; for the Italian SPEECON corpus, it is 83 ms. (For the Italian den Os corpus, it cannot be computed from the available data.) If English had more very short stressed syllables than German and Italian, one might expect the standard deviation to be higher (given that the mean duration is almost the same in the three languages). While the standard deviation for English is somewhat higher than Italian, the values for English and German are almost equal. However, this is not conclusive regarding the frequency of short syllables in the three languages.

We further investigated the data by categorizing the syllables into durational ranges of 50 ms. For English, German, and the Italian SPEECON corpus, this is straightforward, as the duration of each syllable token is given (Figure 6A). (For the den Os Italian corpus, this information is not available.) Using this method, it can be seen that the distributions for the three languages are in fact significantly different (see Table 4). The percentage of syllables whose durations are less than 100 ms is 8.1% in English, 2.8% in German, and only 1.4% in Italian. It does appear, then, that British English has a higher frequency of very short stressed syllables than German or Italian. Defining syllables with durations below 100 ms as "b100" syllables, we performed chi-square tests to see if the proportion of b100 syllables in English was significantly higher than in German and Italian; in both cases, the difference was highly significant (English vs. German, $\chi^2 = 1913.84$, $p < .0001$; English vs. Italian, $\chi^2 = 99.20$, $p < .0001$).

While the den Os Italian corpus does not provide explicit data about the distribution of syllable durations, we do have the count, mean duration, and standard deviation for each of the phoneme categories in Table 3. Let us assume that durations within each phoneme category are normally distributed, as shown in Figure 6B.[4] In that case, given the mean and standard deviation of each category, we can estimate the proportion (and hence the count) of tokens in each category whose duration is below a certain level. We can then estimate the overall count for syllables in different durational ranges. By this method, the proportion of stressed syllables in the den Os corpus with durations below 100 ms is just 0.4%, lower than our English and German corpora and even lower than the SPEECON Italian corpus (compare Figures 6A and 6B). Thus, both the SPEECON and den Os corpora suggest that the frequency of very short stressed syllables is lower in Italian than in British English.

At a typical moderate tempo of quarter-note = 100, a quarter-note is 600 ms. In that case, a stressed syllable whose natural spoken duration is less than 100 ms would, perhaps, sound awkward and unnatural if set to an eighth-note (of 300 ms) or quarter-note (600 ms); a 16th-note (150 ms) would offer a closer approximation to the syllable's natural length. Since the frequency of such syllables is significantly higher in English than in German or Italian, it seems reasonable to propose this as an explanation for the much higher frequency of the SS pattern in English songs.

To explore the data further, we examined the durations of syllables with different vowels. As noted earlier, we had observed informally that many instances of the SS appear to occur with short British English vowels: [æ] (as in *pat*), [ɛ] (as in *pet*), [ɪ] (as in *pit*), [ɒ] (as in *cot*), [ʌ] (as in *cut*), and [ʊ] (as in *put*). Italian has no short vowels of any kind. German does not have [æ] or [ʌ], but the other four short vowels of British English (or very similar counterparts) are all found in German (Hall, 1992). Thus, we examined the mean length of stressed syllables with each of these vowels, in English and German. The data are shown in Table 5. (The table also shows short vowels that are only present in one language or the other.) For all four of the vowels present in both languages, English syllables are indeed shorter than German ones. In three of the four cases, the difference is statistically significant: For [ɛ], $t(2545) = 11.28$, $p < .0001$; for [ʊ] $t(612) = 4.48$, $p < .0001$; for [ɒ], $t(1786) = 2.47$, $p < .05$; for [ɪ], $t(1952) = 1.39$, *ns*. The table also shows the proportion of "b100" syllables in each category—stressed syllables that are less than 100 ms in duration. For all four of the shared vowels, the proportion of b100 syllables in English is higher than the overall proportion in the corpus (8.1%). These four vowels account for 61% of the b100 syllable tokens in the English corpus (but only 33% of stressed syllable tokens overall). To a large extent, then, these vowels account for the higher

[4] Admittedly the assumption of a normal distribution is problematic. Syllable durations are not necessarily normally distributed; the German and Italian data in Figure 6A look close to normal but the English data are less so. Note, however, that our assumption only applies to the phoneme categories in Table 3, not to the distribution as a whole.

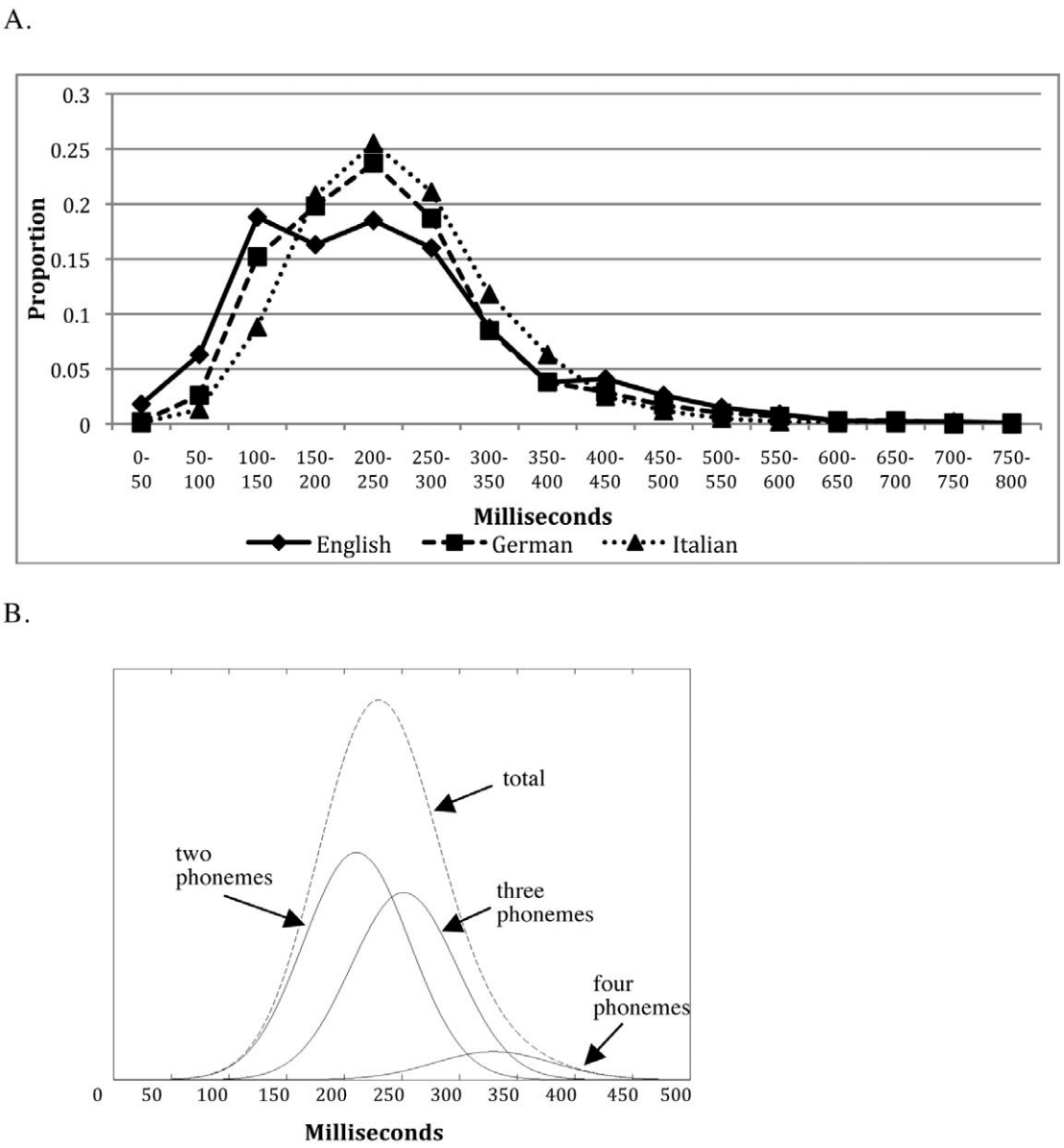


FIGURE 6. (A) Distribution of stressed syllables in the English, German, and SPEECON Italian corpora by duration. (B) Estimated distribution of stressed syllables in the den Os Italian corpus. Solid lines show the distributions for the categories in Table 3 (syllables with two through four phonemes), assuming a normal distribution for each category; dotted line shows overall distribution.

proportion of very short stressed syllables in English than in German. Returning to our English musical data, we counted the number of instances of the SS using each vowel sound. For this, we used a somewhat larger corpus: 274 songs from *The Songs of England* (see Table 1), containing 184 occurrences of the SS. We found that 86% of SS tokens in the corpus use one of the six short vowels of British English; 64% of SS tokens use one of the four short vowels shared with German.

Thus, while the long/short vowel distinction does not completely explain the higher frequency of the SS in English songs than in German or Italian ones, we find that it is relevant to the SS phenomenon. Certain theoretically "short" vowel sounds are, indeed, particularly short, especially in British English. And these vowel sounds account for a large proportion of occurrences of the SS.

DISCUSSION

Our investigations show that the Scotch Snap pattern is widely prevalent in English-language songs in general, but rare or absent in German and Italian songs. In English and Scottish songs it is often associated with

TABLE 4. Short Stressed Syllables in English, German, and Italian.

	% of tokens with duration less than			
	50 ms	100 ms	150 ms	200 ms
English	1.8	8.1	26.9	43.2
German	0.2	2.8	18.0	37.8
Italian (SPEECON)	0.1	1.4	10.2	31.0
Italian (den Os, estimated)	0.0	0.4	5.2	26.2

short vowels. This type of vowel does not exist in Italian speech, which offers a partial explanation for the absence of SS's in Italian songs. German speech, however, does have short vowels, so a further explanation is needed for the fact that the SS is common in British songs but not in German ones.

We have found such an explanation by analyzing speech corpora in the three languages. There is a significantly higher frequency of very short stressed syllables in British English than in either of the other languages; the proportion of stressed syllables shorter than 100 ms is much higher in English than in German or Italian. These very short syllables are similar in duration to a 16th note at a normal tempo (if not shorter). It is desirable to shorten the notes that accompany such syllables, and hence to replace the most typical division of a quarter-note tactus (two equal notes) with an SS. This accounts for the frequent occurrence of SS's in songs with English texts. In German and Italian songs, there is less reason for such adjustments, since the natural duration of a stressed syllable in those languages is rarely far below the normal duration of an eighth-note.

Further investigations showed that four short vowel sounds—[ɪ], [ɛ], [ʊ], and [ɒ]—account for more than 60% of the very short (less than 100 ms) stressed syllable tokens in British English. For each of these vowels, the mean length of stressed syllables is lower in English than in German; in three of the four cases, the difference is statistically significant. And these four vowels account for 64% of the SS tokens in our English musical corpus. Thus, the SS phenomenon appears to be due largely to a few vowels that tend to be especially short in British English.

Our argument proposes a causal link between speech and music. It is true that the speech data used in our study is taken from modern corpora, while most of the music was written at least 100 years ago (in some cases, much earlier than that). Languages change over time, of course, and it is likely that the speech rhythm of English and other languages has evolved quite significantly since the 19th century (unfortunately, no quantitative data is available

TABLE 5. Syllables with Short Vowels in the English and German Corpora*.

Vowel (with English and German sample words)	English		German	
	mean syllable duration	% < 100 ms	mean syllable duration	% < 100 ms
ɛ (*pet, recht*)	206	15.4	266	0.6
ɪ (*pit, bin*)	212	13.5	220	2.5
ɒ (*cot, voll*)	207	14.4	220	0.8
ʊ (*put, Grund*)	175	19.8	239	2.3
æ (*cat*, –)	217	6.2	—	
ʌ (*cut*, –)	217	7.8	—	
a (–, *Hand*)	—		223	2.4
ʏ (–, *fünf*)	—		247	1.6
œ (–, *Köln*)	—		253	0.8

*An empty cell indicates that the vowel is not used in that language.

to examine this issue). However, our results—in particular, the fact that the vowels that tend to be very short in modern British speech account for a large proportion of occurrences of the SS—suggest that our speech data is representative, in the relevant ways, of the speech of the composers of our musical corpora, and the singers for whom they were writing.

If our argument is correct, the SS is a modification of the common European rhythmic idiom to cater for a specific feature of British speech. This does not necessarily imply that every SS is directly caused by such constraints. Over time it became an accepted part of the style of certain kinds of English-language song, and more specifically, of Scottish songs. This may well be the reason why Quantz and others, when they heard a Lombardic rhythm, associated it with Scottish music, and then Scottish singers, composers, and editors in turn were encouraged to use SS's in their songs, even, sometimes, when the stressed note fell on a long vowel. All this is especially likely to have occurred during the Romantic period, when the more conspicuous characteristics of national folk song from countries outside the center of Western Europe—whether Scottish, Hungarian, Polish, Russian, Spanish, or Turkish—were popular in fashionable circles and were widely exploited by composers.

On the other hand, while German composers may have experienced the same tendency to shorten a note carrying a stressed short vowel, they would have felt it in a much weaker form, which (at least in the corpus we examined) was apparently not sufficient to motivate any departure from the common musical style. It has often been noted that German vocal music was profoundly influenced by Italian style in the 17th, 18th and 19th centuries, so that any slight inclination towards the SS

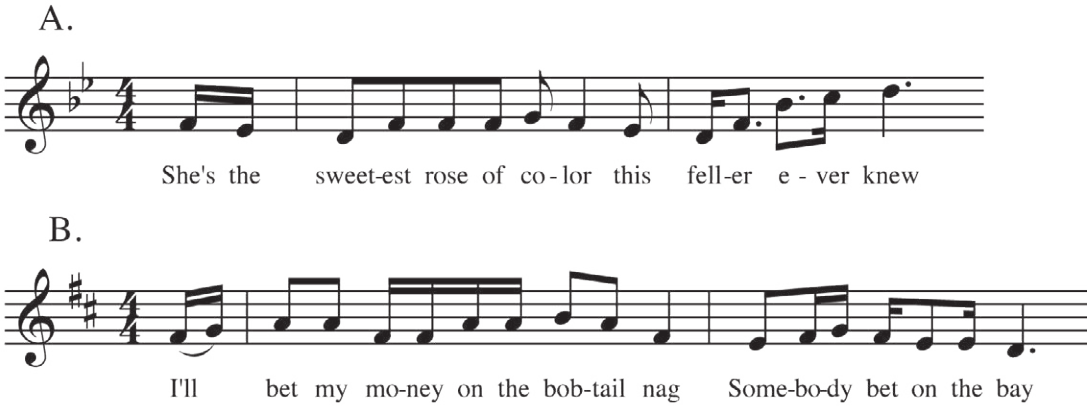


FIGURE 7. Excerpts from two 19th-century American songs (as transcribed in Fox & Weissman, 2007). (A) "The Yellow Rose of Texas" (traditional, first published 1858); (B) Stephen Foster, "Camptown Races" (1850).

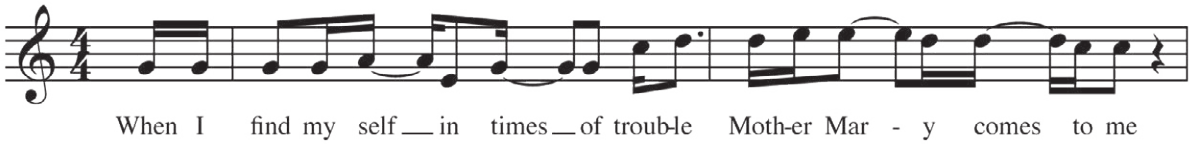


FIGURE 8. The Beatles, "Let it Be."

may have been overridden by a strong impulse to imitate the smoother rhythms of Italian song.

Several directions for future work come to mind. As noted earlier, our musical corpora included only songs in simple meters (i.e., those with the tactus divided in two—2/4, 3/4, and 4/4), not compound meters (with the tactus divided in three, such as 6/8). If our argument is correct, we might expect that SS-like patterns in compound meter (for example, an eighth-note on the beat followed by a quarter-note in 6/8 time) would also be more common in English music than in German or Italian; this would be worth investigating.

Another natural extension of this work would be to examine the correlation between SS frequency and syllable duration in other languages and dialects. It would be of interest to compare Scottish and English speech with regard to vowel duration, given that the SS pattern is more frequent in Scottish songs than in English ones, but (as noted earlier) suitable data for Scottish speech is not currently available. Also of interest in this regard is American English. Informal observation suggests that the SS pattern is quite common in American vocal music. A major complicating factor here, however, is the influence of African-American music. Consider the examples in Figure 7, from two well-known 19th-century American songs. The first contains a straightforward case of the SS, on "feller." The second one contains a pattern similar to the SS, on the words "bet on the"; the pattern features a strong-beat sixteenth-note, but in this case it is followed by an eighth and another sixteenth (rather than a dotted eighth). This pattern of sixteenth-eighth-sixteenth is usually described as a syncopation, resulting from African-American influence (Berlin, 2001). But if this is the case, it seems likely that the very similar SS pattern—in cases such as Figure 7A—was due partly to African-American influence as well, or at least, was reinforced by that influence. In the case of American music, then, explaining the SS as arising from the speech rhythm of American English becomes less plausible.

The issue becomes even more complex in the 20th century, when the syncopations of African-American music come to permeate not only much American popular music but that of other nations as well. Consider Figure 8, from the Beatles' "Let it Be." The passage appears to contain an SS pattern, on the word "trouble." But it also contains a number of other rhythmic figures that would normally be considered syncopations. In each of these syncopations, a syllable is shifted (from what seems to be its "correct" position) one eighth or sixteenth beat to the left: for example, the rhythm of "myself" can be analyzed as an underlying rhythm of two eighths with the second note shifted one sixteenth-beat earlier. But the same reasoning could also be applied to "trouble"; the pattern could be explained as a two-eighth-note pattern with a syncopated second note. Again, it is difficult to say whether apparent SS's such as this one are related to the SS pattern of earlier Anglo-American music, or whether they simply represent the more general phenomenon of anticipatory syncopation.

Within the chronological and geographical limits of this study, however, we believe that we have supplied convincing evidence that the difference in SS frequency between certain European musical styles can be attributed to differences between the corresponding languages. Taken together with other recent research (Huron & Ollen, 2003; Patel & Daniele, 2003), our study therefore provides additional evidence for the general idea that linguistic rhythm is an important factor in the shaping of musical rhythm. We hope that this study will stimulate inquiry into other cultures where specific characteristics of language and speech can explain distinctive features of national musical style.

Author Note

We are grateful to Florian Schiel for assistance with the German speech data, to David Reibel for advice on various subjects, and to Els den Os and Francesco Cutugno for assistance with the Italian speech data.

Correspondence concerning this article should be addressed to Nicholas Temperley, 805 W. Indiana, Urbana, IL 61801. E-mail: ntemp@illinois.edu

References

Abete, G., Cutugno, F., Ludusan, B., & Origlia, A. (2010). Pitch behavior detection for automatic prominence recognition. *Proceedings of Workshop on Prosodic Prominence: Perceptual and Automatic Identification.* Retrieved from http://speechprosody2010.illinois.edu/papers/102001.pdf

Abraham, G. (1974). *The tradition of Western music.* Berkeley, CA: University of California Press.

Auran C., Bouzon, C., & Hirst, D. (2004). The Aix–MARSEC project: An evolutive database of spoken British English. *Proceedings of Speech Prosody.* Retrieved from http://aune.lpl.univ-aix.fr/~EPGA/marsec_com/Auran_Bouzon_Hirst_SP2004.pdf

Baroni, M. (1991). Interrelationships between verbal and musical grammars. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, Language, Speech and Brain: Proceedings of an International Symposium at the Wenner-Gren Center* (pp. 419–428). Basingstoke, UK: Macmillan.

Berlin, E. A. (2001). Ragtime, (2) History. In S. Sadie (Ed.), *The new Grove dictionary of music* (Vol. 20, pp. 756–759). London, UK: Macmillan.

Burger, S., Weilhammer, K., Schiel, F., & Tillmann, H.G. (2000). Verbmobil data collection and annotation. In W. Wahlster (Ed.), *Foundations of speech-to-speech translation* (pp. 537–549). Berlin: Springer.

den Os, E. (1988). *Rhythm and tempo of Dutch and Italian: A contrastive study.* Utrecht, The Netherlands: Elinkwijk BV.

Elliott, K., & Shire, H. M. (Eds.). (1975). *Music of Scotland 1500–1700.* Musica Britannica (Vol. 15). London, UK: Royal Musical Association.

Fant, G., Kruckenberg, A., & Nord, L. (1991). Stress patterns and rhythm in the reading of prose and poetry with analogies to music performance. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, Language, Speech and Brain: Proceedings of an International Symposium at the Wenner-Gren Center* (pp. 380–407). Basingstoke, UK: Macmillan.

Fiske, R. (1983). *Scotland in music.* Cambridge, UK: Cambridge University Press.

Fox, D., & Weissman, D. (2007). *The great family songbook.* New York: Black Dog & Leventhal.

Giegerich, J. (1992). *English phonology: An introduction.* Cambridge, UK: Cambridge University Press.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 515–546). Berlin, Germany: Mouton de Gruter.

Hall, C. (1992). *Modern German pronunciation: An introduction for speakers of English.* Manchester, UK: Manchester University Press.

Halle, J., & Lerdahl, F. (1993). A generative text–setting model. *Current Musicology*, *55,* 3–23.

Harris, J. (1994). *English sound structure.* Oxford, UK: Blackwell.

Heffner, R.-M. S. (1950). *General phonetics.* Madison, WI: University of Wisconsin Press.

Holman, P. (2001). Lampe, John Frederick. In S. Sadie (Ed.), *The new Grove dictionary of music* (Vol. 14, pp. 199–202). London, UK: Macmillan.

Houle, G. (1987). *Meter in music, 1600–1800.* Bloomington, IN: Indiana University Press.

Huron, D., & Ollen, J. (2003). Agogic contrast in French and English themes: Further support for Patel and Daniele (2003). *Music Perception, 21,* 267–271.

Johnson, D. (2001). Scotch snap. In S. Sadie (Ed.), *The new Grove dictionary of music* (Vol. 22, p. 905). London, UK: Macmillan.

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics.* Boston, MA: Wadsworth.

Lass, R. (1984). *Phonology: An introduction to basic concepts.* Cambridge, UK: Cambridge University Press.

Latham, A. (Ed.). (2002). Scotch snap [Scotch catch] (unsigned). *The Oxford companion to music* (pp. 1131–1132). Oxford, UK: Oxford University Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Messinger, H. (1973). *Langenscheidt's concise German dictionary*. London, UK: Hodder and Stoughton.

Palmer, C., & Kelly, M. H. (1992). Linguistic prosody and musical meter in song. *Journal of Memory and Language*, *31,* 525–542.

Patel, A. (2008). *Music, language, and the brain.* Oxford, UK: Oxford University Press

Patel, A., & Daniele, J. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87*, B35–B45.

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *Journal of the Acoustical Society of America, 199,* 3034–3037.

Quantz, J. (1752). *Versuch einer Anweisung die Flöte traversiere zu spielen.* [An attempt at an instruction manual for playing the transverse flute] Berlin, Germany: Johann Friedrich Voss.

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In B. Bell & I. Marlien (Eds.), *Proceedings of Speech Prosody, Aix-en-Provence* (pp. 115–120). Aix-en-Provence, France: Laboratoire Parole et Langage.

Rebora, P. (1958). *Cassell's Italian dictionary.* New York: Funk & Wagnalls.

Siemund, R., Hoege, H., Kunzmann, S., & Marasek, K. (2000). SPEECON—Speech data for consumer devices. *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC).* Retrieved from http://www.speech-dat.org/speecon/public_docs/lrec2000.pdf

Sundberg, J., Nord, L., & Carlson, R. (Eds.). (1991). *Music, Language, Speech and Brain: Proceedings of an International Symposium at the Wenner-Gren Center.* Basingstoke, UK: Macmillan.

Temperley, D. (2010). Modeling common-practice rhythm. *Music Perception*, *27*, 355–376.

Temperley, N. (1981). The Old Way of Singing: Its origins and development. *Journal of the American Musicological Society*, *34*, 511–44.

VanHandel, L., & Song, T. (2010). The role of meter in compositional style in 19th century French and German art song. *Journal of New Music Research, 39*, 1–11.