Comparing the rhythm and melody of speech and music: The case of British English and French

Aniruddh D. Patel,[a] John R. Iversen, and Jason C. Rosenberg
The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, California 92121

(Received 8 September 2005; revised 1 February 2006; accepted 2 February 2006)

For over half a century, musicologists and linguists have suggested that the prosody of a culture's native language is reflected in the rhythms and melodies of its instrumental music. Testing this idea requires quantitative methods for comparing musical and spoken rhythm and melody. This study applies such methods to the speech and music of England and France. The results reveal that music reflects patterns of durational contrast between successive vowels in spoken sentences, as well as patterns of pitch interval variability in speech. The methods presented here are suitable for studying speech-music relations in a broad range of cultures. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2179657]

PACS number(s): 43.70.Fq, 43.71.Es, 43.71.An, 43.75.Cd, 43.70.Kv [BHS] Pages: 3034–3047

I. INTRODUCTION

A. Aims

Humans produce organized rhythmic and melodic patterns in two forms: prosody and music. While these patterns are typically studied by different research communities, their relationship has long interested scholars from both fields. For example, linguists have borrowed musicological concepts in building prosodic theories (Liberman, 1975; Selkirk, 1984), and musicologists have used tools from linguistic theory to describe musical structure (Lerdahl and Jackendoff, 1983). Despite this contact at the theoretical level, there has been remarkably little empirical work comparing rhythmic or melodic structure across domains. There are reasons to believe such work is warranted. One such reason, which motivates the current study, is the claim that a composer's music reflects prosodic patterns in his or her native language. This idea has been voiced repeatedly by scholars over the past half century. For example, the English musicologist Gerald Abraham explored this topic at length (1974), citing as one example Ralph Kirkpatrick's comment on French keyboard music:

"Both Couperin and Rameau, like Fauré and Debussy, are thoroughly conditioned by the nuances and inflections of spoken French. On no Western music has the influence of language been stronger." (p. 83)

Kirkpatrick (a harpsichordist and music scholar) was claiming that French keyboard music sounded like the French language. Similar claims have been made about the instrumental music of other cultures. The linguist Hall, for example, suggested a resemblance between Elgar's music and the intonation of British speech (Hall, 1953). What makes these claims interesting is that they concern *instrumental* music. It might not be surprising if vocal music reflected speech prosody; after all, such music must adapt itself to the rhythmic and melodic properties of a text. In contrast, the notion that speech patterns are mirrored in instrumental music is much more controversial.

While provocative, until recently this idea had not been systematically tested, likely due to a lack of methods for quantifying prosody in a way that could be directly compared to music. Patel and Daniele (2003) sought to overcome this problem (with regard to rhythm) by using a recently-developed measure of temporal patterning in speech, the normalized pairwise variability index or nPVI. The nPVI measures the degree of durational contrast between successive elements in a sequence, and was developed to explore rhythmic differences between "stress-timed" and "syllable-timed" languages (Low, 1998; Low *et al.*, 2000). Empirical work in phonetics has revealed that the nPVI of vowel durations in sentences is significantly higher in stress-timed languages such as British English than in syllable-timed languages such as French (Grabe and Low, 2002; Ramus, 2002). The reason for this is thought to be the greater degree of vowel reduction in the former languages (Dauer, 1983, 1987; Nespor, 1990). Patel and Daniele applied the nPVI to the durations of notes in instrumental classical themes from England and France, and found that English music had a significantly higher nPVI than French music. This earlier work illustrates a cross-cultural approach to comparing prosodic and musical structure which is extended in the current study. This approach is based on determining whether quantitative prosodic differences between languages are reflected in music (cf. Wenk, 1987).

While Patel and Daniele (2003) focused on composers from the turn of the 20th century (a time of musical nationalism), subsequent work showed that their finding generalized to a much broader historical sample of composers from England and France (Huron and Ollen, 2003). Furthermore, it appears that other European cultures with stress-timed languages tend to have higher musical nPVI values than cultures with syllable-time languages (Huron and Ollen, 2003), though interesting exceptions exist (Patel and Daniele, 2003b; Daniele and Patel, 2004). These studies indicate that prosody and instrumental music can be meaningfully compared using quantitative methods. They also raise two key

[a]Author to whom correspondence should be addressed. Tel: 858-626-2085; Fax: 858-626-2099; electronic mail: apatel@nsi.edu

 0001-4966/2006/119(5)/3034/14/$22.50

questions which are the focus of the current study, as detailed in Secs. I A 1 and I A 2 below. As in the previous work, the current study focuses on British English and continental French (henceforth English and French). One of the principal goals, however, is to address issues and develop methods of broad applicability to speech-music research.

1. Are differences in durational contrast a byproduct of variability differences?

Although the nPVI's name refers to a "variability index," it is in fact a measure of durational contrast. This is evident from the nPVI equation,

$$\mathrm{nPVI} = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right|, \quad (1)$$

where m is the number of durational elements in a sequence d_k is the duration of the kth element. The nPVI computes the absolute difference between each successive pair of durations in a sequence, normalized by the mean duration of the pair. This converts a sequence of m durations (e.g., vowel durations in a sentence) to a sequence of $m-1$ contrastiveness scores. Each of these scores ranges between 0 (when the two durations are identical) and 2 (for maximum durational contrast, i.e., when one of the durations approaches 0). The mean of these scores, multiplied by 100, yields the nPVI of the sequence. The nPVI is thus a contrastiveness index and is quite distinct from measures of overall variability (such as the standard deviation) which are insensitive to the order of observations in a sequence. Indeed, one cannot compute the nPVI of a given sequence from its standard deviation or vice versa. Nevertheless, at the population level differences in overall variability of two sets of sequences will inevitably drive some degree of nPVI difference between the sets, simply because sequences with greater variability are likely to contain neighbors of greater durational contrast (cf. Sadakata and Desain, submitted).

This point is relevant to the comparison of English and French because there are reasons to expect that vowels in English sentences should exhibit higher overall durational variability than vowels in French sentences. One such reason is that vowel duration in English is substantially modulated by stress and vowel reduction, factors which play much less of a role in modulating vowel duration in French (Delattre, 1966; 1969). Hence it is important to know if linguistic nPVI differences between the languages are simply a by-product of variability differences. A similar question applies to music. Should this be the case, then music may simply reflect differences in linguistic temporal variability, with the nPVI difference being a by-product of such differences.

To examine these issues, a measure of overall variability for each sentence and musical theme is computed in this study in order to examine the relationship between variability and nPVI. Specifically, a Monte Carlo method is used to quantify the likelihood of observing an nPVI difference of a given magnitude between two languages (or two musics) given existing differences in variability.

2. Is speech melody reflected in music?

The original intuition of a link between prosody and instrumental music was not confined to rhythm, but encompassed melody as well (e.g., Hall, 1953). The current study addresses this issue via a quantitative comparison of intonation and musical melody. Earlier comparative work on rhythm had the benefit of an empirical measure which could readily be applied to music (the nPVI). In the case of intonation, no such measure was available. To overcome this problem, this study employs a recent computational model of speech intonation perception known as the "prosogram" (Mertens, 2004a, 2004b). The prosogram converts a sentence's fundamental frequency (Fo) contour into a sequence of discrete tonal segments, producing a representation which is meant to capture the perceived pitch pattern of a speech melody. This representation allows a quantitative comparison of pitch variability in speech and music. Further details on the prosogram and measures of variability are given in the next section.

B. Background

1. Rhythm

Speech rhythm refers to the way languages are organized in time. Linguists have long held that certain languages (such as English and French) have decidedly different rhythms, though the physical basis of this difference has been hard to define. Early ideas that the difference lay in the unit produced isochronously (either stresses or syllables) have not been supported by empirical research (e.g., Roach, 1982, Dauer, 1983). Some linguists have nevertheless retained the "stress-timed" vs "syllable-timed" terminology, likely reflecting an intuition that languages placed in these categories do have salient rhythmic differences (Beckman, 1992). Examples of languages placed in these categories are British English, Dutch, and Thai (stress-timed) vs French, Spanish, and Singapore English (syllable timed) (Grabe and Low, 2002).

Recent years have seen the discovery of systematic temporal differences between stress-timed and syllable-timed languages (e.g. Ramus *et al.*, 1999; Low *et al.*, 2000). These discoveries illustrate the fact that "rhythm" in speech should not be equated with isochrony. That is, while isochrony defines one kind of rhythm, the absence of isochrony is not the same as the absence of rhythm. Rhythm is the systematic temporal and accentual patterning of sound. Languages can have rhythmic differences which have nothing to do with isochrony. For example, Ramus and co-workers (1999) demonstrated that sentences in stress-timed (vs syllable-timed) languages had greater durational variability in "consonantal intervals" (consonants or sequences of abutting consonants, regardless of syllable or word boundaries) and a lower overall percentage of sentence duration devoted to vowels. (They termed these two measures ΔC and $\%V$, respectively). These differences likely reflect phonological factors such as the greater variety of syllable types and the greater degree of vowel reduction in stress-timed languages (Dauer, 1983). Another difference between stress-timed and syllable-timed languages, noted previously, is that the durational contrast

between adjacent vowels in sentences is higher in the former types of languages (as measured by the nPVI), probably also due to the greater degree of vowel reduction in these languages (Low *et al.*, 2000; Grabe and Low, 2002; Ramus, 2002).[1] An interesting point raised by these recent empirical studies is that languages may fall along a rhythmic continuum rather than forming discrete rhythm classes (cf. Nespor, 1990; Grabe and Low, 2002). This "category vs continuum" debate in speech rhythm research has yet to be resolved, and is largely orthogonal to the issues addressed here.

Of the different temporal measures described above, vowel-based measures of rhythm are the most easily transferred to music research. This is because musical notes can roughly be compared to syllables, and vowels form the core of syllables. It therefore seems plausible to compare vowel-based rhythmic measures of speech to note-based rhythmic measures of music. Of the two vowel-based measures discussed above (*%V* and nPVI), the latter can be sensibly applied to music by measuring the durational contrast between successive notes in a sequence. This approach is taken in the current work.

Since the focus here is on English and French, it is worth asking about the robustness of the nPVI difference between these two languages. Significant nPVI differences between English and French have been reported by four published studies, one based on vowel durations in spontaneous speech (Grabe *et al.*, 1999), and three based on vocalic interval durations in read speech (Grabe and Low, 2002; Ramus, 2002; Lee and Todd, 2004). (A vocalic interval is defined as the temporal interval between a vowel onset and the onset of the next consonant in the sentence; a vocalic interval may thus contain more than one vowel and can span a syllable or word boundary, cf. Ramus *et al.*, 1999; Grabe and Low, 2002. The choice of vowels vs vocalic intervals makes little difference when comparing the nPVI of English and French, cf. Secs. II A and II B.) One notable finding is that nPVI values for English and French vary considerably from study to study. For example, Ramus (2002) reported values of 67.0 and 49.3 for English and French, respectively, while Lee and Todd (2004) reported values of 83.9 and 54.3. Both studies measured vocalic intervals in read speech. Possible sources of this discrepancy include differences in speech materials, speech rate, and the criteria for the placement of boundaries between vowels and consonants (cf. Dellwo *et al.*, 2006). Further research is needed to clarify this issue. What one can say confidently, however, is that the finding that English has a significantly higher nPVI than French (within a given study) appears highly robust.[2]

2. Melody

A central issue for comparing melody in speech and music is how to represent speech melodies. Several choices exist. One choice is to use the raw Fo contours of sentences. Another choice is to use sequences of abstract phonological tones, such as high (*H*) and low (*L*) tones in autosegmental-metrical theories of intonation (e.g., Pierrehumbert, 1980; Ladd 1996). This study opts for a representation that is neither as detailed as raw Fo contours nor as abstract as autosegmental-metrical approaches. This is the "prosogram" representation of intonation (Mertens, 2004a, 2004b; cf. d'Alessandro and Mertens, 1995).

The prosogram aims to provide a representation of intonation as perceived by human listeners, and thus follows in the tradition of Fo stylization based on perceptual principles (Rossi, 1971; Rossi, 1978a; Rossi, 1978b; 'tHart *et al.*, 1990). It is based on empirical research which suggests that pitch perception in speech is subject to four perceptual transformations. The first is the segregation of the Fo contour into syllable-sized units due to the rapid spectral and amplitude fluctuations in the speech signal (House, 1990). The second is a threshold for the detection of pitch movement within a syllable (the "glissando threshold"). The third, which applies when pitch movement is detected, is a threshold for detection of a change in the slope of a pitch movement within a syllable (the "differential glissando threshold"). The fourth, which applies when the pitch movement is subliminal, is temporal integration of Fo within a syllable (d'Alessandro and Castellengo, 1994; d'Alessandro and Mertens, 1995). The prosogram instantiates the latter three transformations via an algorithm which operates on the vocalic nuclei of syllables (phonetic segmentation is provided by the user). As a result of these transforms, the original Fo contour of a sentence is converted to a sequence of discrete tonal segments. An example of the model's output is given in Fig. 1, which shows the original Fo contour (top) and the prosogram (bottom) for the English sentence "Having a big car is not something I would recommend in this city." Figure 1 reveals why the prosogram is useful to those interested in comparing speech and music. The representation produced by the prosogram is quite musiclike, consisting mostly of level pitches. (Some syllables are assigned pitch glides if the amount of Fo change is large enough to exceed a perceptual threshold.) On a cognitive level, this is interesting because it implies that the auditory image of speech intonation in a listener's brain has more in common with music than has been generally appreciated. On a practical level, the dominance of level pitches means that intonation patterns in different languages can be compared using tools that can also be applied to music, e.g., statistical measurements of pitch height or pitch interval patterns.[3]

The current study uses the prosogram to examine a simple aspect of the statistical patterning of spoken intonation, namely pitch variability. Specifically, pitch variability in English and French speech is quantified from prosograms and compared to pitch variability in English and French musical themes. Prior studies of pitch variability in the two languages have produced contradictory results. Maidment (1976) computed running Fo from a laryngograph while speakers ($n=16$) read a $2\frac{1}{2}$ min passage of prose. He reported the mean and standard deviation of the Fo contours produced by each speaker. When converted into the coefficient of variation (standard deviation/mean), a significantly higher degree of variability is found for the English than for the French speakers. In contrast, Lee and Todd (2004) reported no significant difference in Fo variability in English and French speech. However, rather than measure raw Fo contours they extracted the Fo at the onset of each vocalic

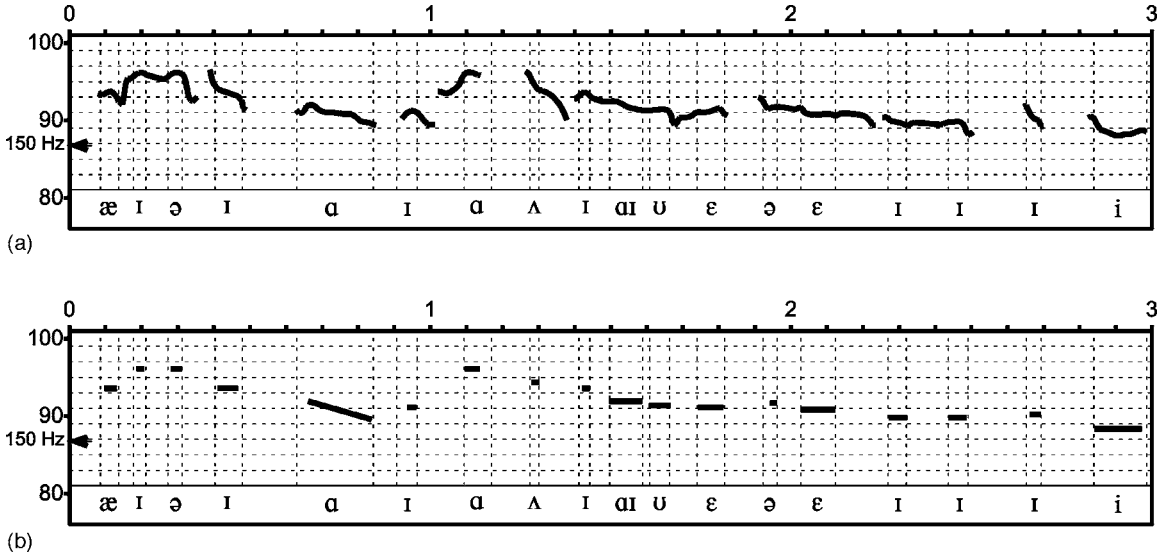

FIG. 1. Illustration of the prosogram, using the British English Sentence "Having a big car is not something I would recommend in this city" as uttered by a female speaker. In both graphs, the horizontal axis along the top shows time in seconds, the vertical axis shows semitones re 1 Hz (an arrow is placed at 150 Hz for reference), and the bottom shows IPA symbols for the vowels in this sentence. The onset and offset of each vowel is indicated by vertical dashed lines above the vowels' IPA symbol. (a) Shows the original Fo contour, while (b) shows the prosogram. In this case, the prosogram has assigned level tones to all vowels save for the vowel in "car," which was assigned a glide. Note that the pitches of the prosogram do not conform to any musical scale.

interval in a sentence and studied the variability of these values across a sentence. (They expressed each value as a semitone distance from the mean vocalic-onset Fo in the utterance, and then computed the standard deviation of this sequence of pitch values.)

These two studies illustrate the fact that pitch variability in speech can be measured in different ways. Whether or not one obtains differences between languages may depend on the method one chooses. For those interested in perception, the prosogram offers a motivated way to study the variability of pitch patterns in speech. Furthermore, it offers two ways to quantify variability, i.e., in terms of pitch height and pitch intervals. The former measures the spread of pitches about a mean pitch (as in Lee and Todd's, 2004 study). The latter measures whether steps between successive pitches tend to be more uniform or more variable in size. Both types of measures were computed in this study in order to compare speech to music.

II. METHODS

A. Corpora

The materials were the same as those in Patel and Daniele (2003). For speech, 20 English and 20 French sentences were taken from the database of Nazzi *et al.* (1998), consisting of four female speakers per language reading five unique sentences each. The sentences had been recorded in a quiet room and digitized at 16 000 Hz. They are short newslike utterances, and have been used in a number of studies of speech rhythm by other researchers (e.g., Nazzi *et al.*, 1998; Ramus *et al.*, 1999; Ramus, 2002). They range from 15 to 20 syllables in length and are approximately 3 s long (see Appendix A for a full list). Table I gives some basic data on sentence characteristics. Sentences contained about 16 vowels on average, most of which were singletons (i.e., a vowel not abutting another vowel). Thus durational computations based on vowels vs vocalic intervals are likely to yield similar results. The original motivation for studying vocalic intervals in studies of speech rhythm was an interest in infant speech perception, under the assumption that infants perform a crude segmentation of the speech stream which only distinguishes between vocalic and nonvocalic (i.e., consonantal) portions (Mehler *et al.*, 1996; Ramus *et al.*, 1999). Since the current work focuses on adult perception of speech, and since vowels are well-established phonological units in language while vocalic intervals are not, this study examines vowels rather than vocalic intervals.

The musical data are themes from turn-of-the 20th century English and French composers, drawn from a musicological sourcebook for instrumental music (*A Dictionary of Musical Themes*, Barlow and Morgenstern, 1983). Themes were analyzed for all English and French composers in the

TABLE I. Some basic statistics on the sentences studied.

	Duration (s) Mean (sd)	Speech rate (syll/s) Mean (sd)	No. Vowels / sentence Mean (sd)	Avg Fo (Hz) Mean (sd)	Total vowels	Singleton vowels
English (n=20)	2.8 (0.2)	5.8 (0.3)	15.7 (1.7)	222.2 (14.1)	314	296
French (n=20)	2.8 (0.2)	6.1 (0.5)	17.3 (1.6)	219.4 (25.7)	346	310

dictionary who were born in the 1800s and died in the 1900s. This era is recognized by musicologists as being a time of musical nationalism, when music is thought to be especially reflective of culture (Grout and Palisca, 2000). It is also not too distant in the past, which is desirable since measurements of speech are based on living speakers and since languages can change phonologically over time. To be included, composers had to have at least five usable themes, i.e., themes that met a number of criteria designed to minimize the influence of language or other external influences on musical structure. For example, themes were excluded if they came from works whose titles suggested a vocal conception or the purposeful evocation of another culture (see Patel and Daniele, 2003 for the full list of criteria). Furthermore, themes were required to have at least 12 notes (to provide a good sample for rhythm measures), and no internal rests, grace notes, or fermatas, which introduced durational uncertainties. These criteria yielded six English composers (137 themes) and ten French composers (181 themes). In reviewing the themes used in the previous study, a few inadvertent errors of inclusion or exclusion were found and corrected, resulting in 136 English and 180 French themes in the current work (see Appendix B for a complete list of composers and themes).

B. Phonetic segmentation

To analyze linguistic nPVI, vowel boundaries were marked in English and French sentences using wide-band speech spectrograms generated with SIGNAL (Engineering Design) running on a modified personal computer (frequency resolution=125 Hz, time resolution=8 ms, one FFT every 3 ms, Hanning window). Vowel onset and offset were defined using standard criteria (Peterson and Lehiste, 1960). Vowel boundaries in this study were marked independently of the boundaries defined by Ramus (2002) for the same set of sentences. Those earlier boundaries, which served as the basis of the nPVI values reported in Patel and Daniele (2003), came from a phonetic segmentation based on a waveform display with interactive playback. In the current study segmentation of the sentences was based on a display showing both the waveform and spectrogram, plus interactive playback. Availability of a spectrogram often resulted in boundary locations which differed from those marked by Ramus.

As noted above, this study focuses on vowels rather than vocalic intervals. However, both quantities were measured and yielded very similar results in the rhythmic analyses. This is not surprising since the great majority of vowels in both languages were singletons (Table I). The results report vowel measurements only (data based on vocalic intervals are available upon request).

C. Duration coding of musical themes

As in earlier work, measurement of duration in musical themes was made directly from music notation. This stands in contrast to the speech measurements, which were (necessarily) based on acoustic signals. Initially this may seem problematic, but as noted in Patel and Daniele (2003) the analysis of acoustic recordings of music raises its own problems, such as which performance of each theme to analyze and how to defend this performance against other available recordings, all of which will differ in the fine nuances of timing. Music notation is thus a reasonable choice since it at least affords an unambiguous record of the composer's intentions. That being said, it is important for future work to study timing patterns in human performances because such performances deviate from the idealized durations of music notation in systematic ways (Repp, 1992; Palmer, 1997). It will be interesting to determine what influence such deviations have on the results of a study like this one.

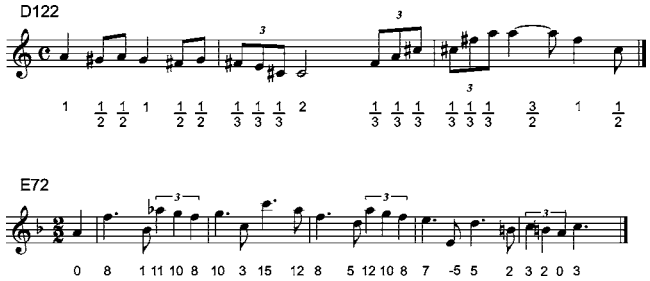

FIG. 2. Examples of duration and pitch coding of musical themes. (D122: Debussy's *Quartet in G minor for Strings, 1st movement, 2nd theme.* E72: Elgar's *Symphony No. 1, in A Flat, Opus 55, 4th movement, 2nd theme.*) D122 illustrates duration coding: the relative duration of each note is shown below the musical staff (see text for details). E72 illustrates pitch coding: each note is assigned a pitch value based on its semitone distance from A4 (440 Hz). The nPVI of note durations in D122 is 42.2. The coefficient of variation (CV) of pitch intervals in E72 is 0.79.

An example of durational coding of a musical theme is shown in Fig. 2 (top). Notes were assigned durations according to the time signature, with the basic beat assigned a duration of 1 (e.g. a quarter note in 4/4 or an eighth note in 3/8), and other notes assigned their relative durations according to standard music notation conventions. Durations were thus quantified in fractions of a beat. Because the nPVI is a relative measure, any scheme which preserves the relative duration of notes will result in the same nPVI for a given sequence. For example, the first note of a theme could always be assigned a duration of 1 and other note durations could be expressed as a fraction or multiple of this value.

D. Computation of rhythmic measures and Monte Carlo analyses

Two rhythmic measures were computed for each sentence and musical theme: the nPVI and the coefficient of variation (CV), the latter defined as the standard deviation divided by the mean. Like the standard deviation, the CV is a measure of overall variability which is insensitive to the order of elements in the sequence. Yet like the nPVI, it is a dimensionless quantity since the same units appear in both the numerator and denominator. It is thus well suited for comparing temporal patterns in speech and music when speech is measured in seconds and music is measured in fractions of a beat. Furthermore, the CV (like the nPVI) is only sensitive to the relative duration of events; scaling these durations up or down by a constant factor does not change its

value. Hence any durational coding scheme for musical themes will produce the same CV as long as the relative durations of events are preserved.

The relationship between nPVI and CV was studied in two ways. First, within each domain (speech or music) linear regressions were performed with CV as the independent variable and nPVI as the dependent variable. This showed the extent to which nPVI was predicted by CV. Next, within each domain a Monte Carlo technique was used to estimate the probability of the observed English-French nPVI difference given existing variability differences. The technique was based on scrambling the order of durations in each sequence, which destroys its temporal structure while retaining its overall variability. For example, if focusing on speech, the sequence of vowel durations in each English and French sentence was randomly scrambled and the nPVI of these scrambled sequences was computed. The difference between the mean nPVI values for the scrambled-English and scrambled-French sentences was recorded. This procedure was repeated 1000 times. The proportion of times that this "scrambled nPVI" difference was equal to or greater than the original nPVI difference was taken as a *p*-value indicating the probability of obtaining an nPVI equal to or greater than the observed nPVI difference between the languages. The same procedure was used for musical themes.

E. Melodic analyses

To quantify pitch patterns in speech, prosograms were computed for all English and French sentences using prosogram version 1.3.6 as instantiated in Praat (Mertens, 2004a; 2004b).[4] As part of the algorithm, an Fo contour was computed for each sentence using the autocorrelation algorithm of Boersma (1993). (Default parameters were used with the exception of frame rate, minimum pitch, and maximum pitch, which were set to 200 Hz, 60 Hz, and 450 Hz, respectively.) Prosogram analysis is based on the vocalic nuclei of syllables. To determine whether a given vowel should be assigned a level tone or a glide, a glide threshold of $0.32/T^2$ semitones/s was used, where T is the duration of a vowel in s. If the rate of pitch change was greater than this threshold, the vowel was assigned a frequency glide (or two abutting glides if the differential glissando threshold was exceeded). The choice of $0.32/T^2$ semitones/s as the glissando threshold is based on perceptual research on the threshold for detecting pitch movement in speech, combined with experiments in which prosogram output is compared to human transcriptions of intonation ('tHart 1976; Mertens, 2004b). Vowels with pitch change below the glide threshold were assigned a level tone equal to their median pitch value. This served as an estimate of the perceived pitch of the syllable, as formerly computed from a time-weighted average of the vowel's Fo contour in earlier versions of the prosogram (e.g., d'Alessandro and Mertens, 1995; Mertens and d'Alessandro, 1995; Mertens *et al.*, 1997).

For maximum comparability to music, only level tones were used in the quantification of pitch variability in speech. Such tones represented 97% of tones assigned to vowels in the current corpus. Occasionally the prosogram did not assign a tonal element to a vowel, e.g., if the intensity of the vowel was too low, the vowel was devoiced (e.g., an unstressed "to" in English being pronounced as an aspirated /t/), or if Praat produced a clearly erroneous Fo value. However, such omissions were rare: 90% of English vowels were assigned level tones and 4% were assigned glides, while 96% of French vowels were assigned level tones and 2% were assigned glides. A typical sentence had about 15 level tones and 1 glide.

Variability in pitch height and of pitch intervals within a sentence was computed via the coefficient of variation (CV). To study pitch height variation, each level tone was assigned a semitone distance from the mean pitch of all level-tones in the sentence, and then the CV of these pitch distances was computed. To study interval variability, adjacent level tones in a sentence were assigned a pitch interval in semitones,

$$\mathrm{st} = 12 \log_2(f2/f1), \tag{2}$$

where $f1$ and $f2$ represent the initial and final tone of the pair, respectively. (Intervals were computed between immediately adjacent level tones only, not between tones separated by a glide.) The CV of these intervals was then quantified. Because the mean appears in the denominator of the CV, measurements of pitch distances and pitch intervals used absolute values in order to avoid cases where mean distance size or mean interval size was equal to or near 0, which would yield mathematically ill-defined CVs. The choice of semitones as units of measurement is based on recent research on the perceptual scaling of intonation (Nolan, 2003). Earlier work by Hermes and Van Gestel (1991) had suggested ERBs should be used in measuring pitch distances in speech. Since the CV is dimensionless, one could measure pitch in speech and music in different units (ERBs vs semitones) and still compare pitch variability across domains using the CV as a common metric. The precise choice of units for speech is unlikely to influence the results reported here.

To quantify melodic patterns in music, musical themes were coded as sequences of pitch values where each value represented a given pitch's semitone distance from A440 (Fig. 2, bottom). This permitted straightforward computation of each tone's semitone distance from the mean pitch of the sequence and of pitch interval patterns (the latter simply being the first-order difference of the pitch values). Measures of pitch height and interval variability were then computed in precisely the same manner as for speech. (Note that the choice of A440 as a referent pitch makes no difference to the measures of variability computed here. Any scheme which preserves the relative position of tones along a semitone scale would yield the same results. For example, one could code each pitch in a musical theme as its distance in semitones from the lowest pitch of the theme, or from the mean pitch of the theme.)

III. RESULTS

A. Rhythm

Table II shows nPVI and CV values for speech and music. Reported *p*-values in this and following tables were computed using the Mann-Whitney U-test, except for *p*-values

TABLE II. nPVI and CV for speech and music (mean and s.e.). The rightmost column gives the probability that the observed nPVI difference is due to the difference in CV.

	English nPVI	French nPVI	p	English CV	French CV	p	$p(\Delta\text{nPVI}\vert\Delta\text{CV})$
Speech (vowels)	55.0 (3.0)	35.9 (1.8)	<0.001	0.55 (0.03)	0.36 (0.02)	<0.001	0.01
Music (notes)	47.1 (1.8)	40.2 (1.9)	<0.01	0.61 (0.02)	0.58 (0.02)	0.34	0.001

associated with Monte Carlo analyses, which were computed as described in Sec. II D. Table II reveals that English and French sentences show a highly significant difference in durational contrastiveness (nPVI) as well as in durational variability (CV). English and French music, on the other hand, show a significant difference in contrastiveness but not in variability.

Regressions of CV on nPVI in each domain are shown in Fig. 3. The linear regressions reveal that within each domain, higher CV is predictive of higher nPVI, though the relationship appears to be stronger in speech than in music (see Fig. 3 captions for regression slopes, r^2 values, and p values).

To assess whether CV differences between the two languages are responsible for linguistic nPVI differences, a Monte Carlo analysis was conducted as described in Sec. II D. The result of this analysis is shown in Fig. 4, which plots the distribution of nPVI differences between English and French speech when the order of vowel durations in each sentence is scrambled and the nPVI difference between the two languages is computed (1000 iterations). The actual nPVI difference (19.1 points) is shown by an arrow. Based on this frequency distribution, the probability of an nPVI difference of 19.1 points or greater is quite small ($p=0.01$). This value is listed in the right-most column of Table II as $p(\Delta\text{nPVI}\vert\Delta\text{CV})$, i.e., the probability of the observed difference in nPVI given the observed difference in variability. A similar analysis was conducted for music, with the resulting p-value being 0.001. Thus it is highly unlikely that variability differences account for nPVI differences in either domain.

B. Melody

Table III shows the results of pitch variability measurements for speech and music. Table III reveals that English and French speech do not differ in the variability of pitch height within utterances, but do show a significant difference in pitch interval size variability, with French having lower

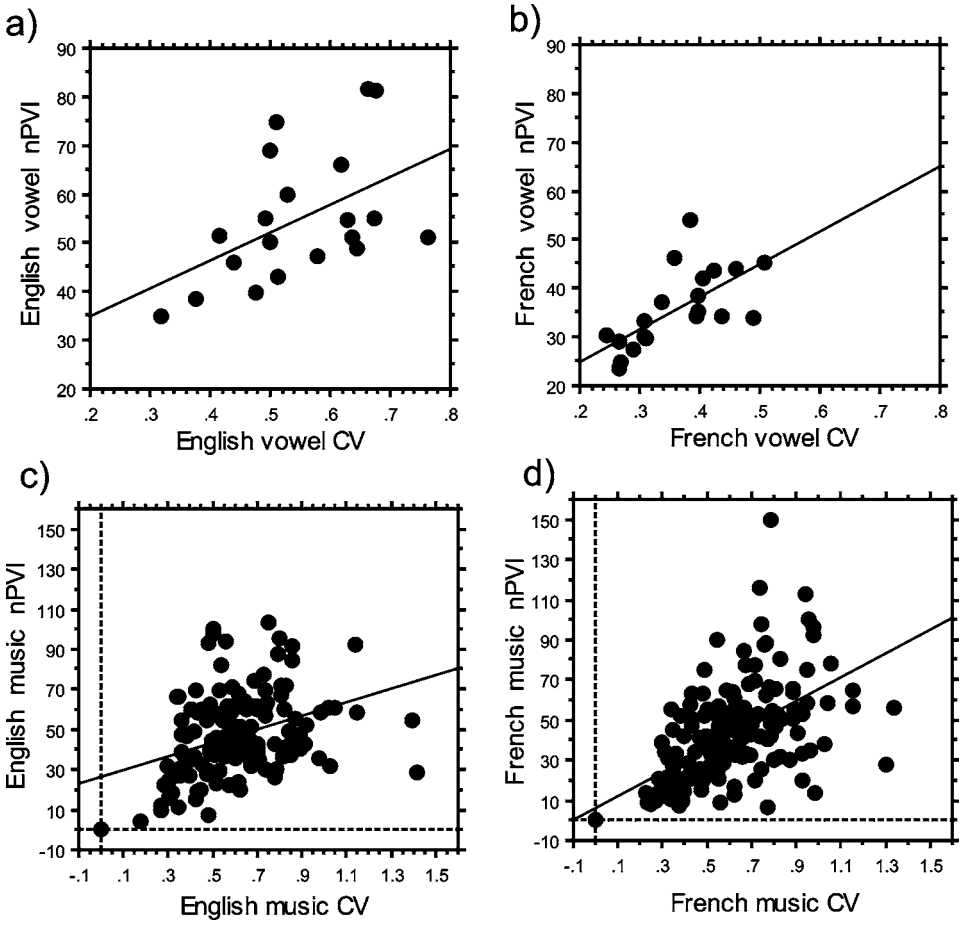

FIG. 3. The relationship between CV and nPVI for speech (a, b) and music (c, d). For speech each dot represents one sentence; for music each dot represents one theme. The best fitting regression line for each panel is also shown. English speech: nPVI=23.5 +57.3×CV, $\mathbf{r^2=0.34}$, df=18, $\mathbf{p=0.03}$; French speech: nPVI=11.7 +66.5×CV, $\mathbf{r^2=0.43}$, df=18, $\mathbf{p<0.01}$; English music: nPVI=26.6 +33.9×CV, $\mathbf{r^2=0.14}$, df=134, $\mathbf{p<0.001}$; French music: nPVI=6.2 +59.1×CV, $\mathbf{r^2=0.36}$, df=178, $\mathbf{p<0.001}$. For the musical data, hatched lines show the lower possible limit of the nPVI and CV at 0 on each axis: the axes range into negative numbers for display purposes only, so that the points at (0,0) can be clearly seen. Themes with a score of 0 for nPVI and CV have notes of a single duration. There were two such English themes and eight such French themes.

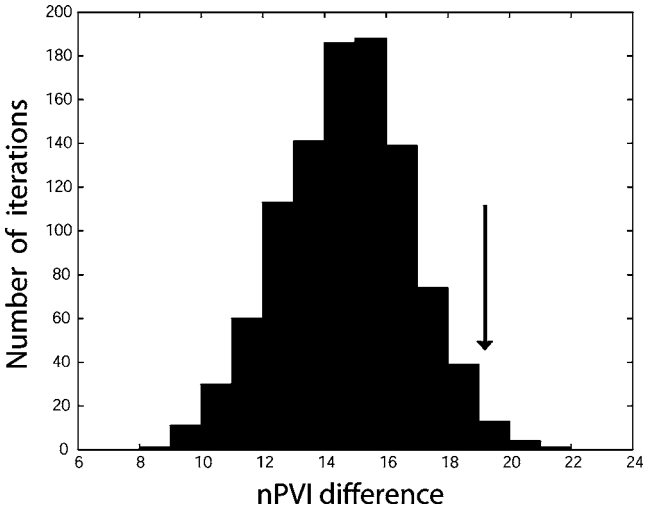

FIG. 4. Result of Monte Carlo analysis for English vs French speech. The actual nPVI difference between the two languages in this study (19.1 points) is shown by an arrow. See text for details.

interval variability than English. With regard to the latter point, it is interesting to note that the average absolute interval size is virtually identical in the two languages (2.1 vs 2.2 st for French vs English, respectively), yet French speech shows significantly lower interval variability than English. In other words, as the voice moves from one vowel to the next the size of the pitch change is more uniform in French than in English speech.

Turning to music, a similar picture emerges: differences in pitch height variability are not significant, but French music has significantly lower pitch interval variability than English music. In other words, as the melody moves from one note to the next the size of the pitch change is more uniform in French than in English music. Once again, this difference exists despite the fact that the average absolute interval size is nearly identical (2.7 vs 2.6 st for French and English music, respectively). Figure 5 shows the data for pitch interval variability in speech and music. Figure 5 shows that the linguistic difference in pitch interval variability between English and French speech is much more pronounced than is the musical difference, a finding reminiscent of earlier work

TABLE III. Pitch variability in speech and music, measured in terms of pitch height (the CV of pitch distances from the mean pitch of a sequence) or pitch intervals (the CV of pitch interval size within a sequence). Mean and s.e. are shown.

	English Pitch CV	French Pitch CV	p
Speech			
Pitch height	0.71 (0.04)	0.75 (0.04)	0.32
Pitch intervals	0.88 (0.05)	0.68 (0.03)	<0.01
Music			
Pitch height	0.69 (0.01)	0.71 (0.01)	0.14
Pitch intervals	0.76 (0.02)	0.71 (0.02)	0.03

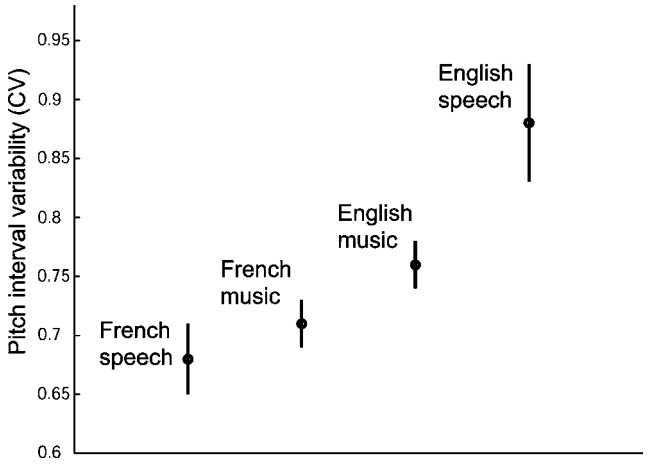

FIG. 5. Pitch interval variability in English and French speech and music. Pitch interval variability is defined as the CV of absolute interval size between pitches in a sequence. Error bars show standard errors.

on the nPVI. This should not be surprising, since music is an artistic endeavor with substantial intracultural variation and (unlike speech) no *a priori* reason to follow melodic norms. What is remarkable is that despite this variation, a significant cultural difference emerges in the same direction as the linguistic difference.

C. Rhythm and melody combined

Having analyzed rhythm and melody independently, it is interesting to combine the results in a single graph showing English and French speech and music in two-dimensional space with rhythm and melody on orthogonal axes. This representation can be referred to as an RM-space plot. Figure 6 shows such a plot, with nPVI being the measure of rhythm and melodic interval variability (MIV) being the measure of melody. MIV is defined as $100 \times$CV of pitch interval variability. Scaling the CV by 100 serves to put MIV in the same general range of values as nPVI. One aspect of this figure deserves immediate comment. Recall from Sec. I B 1 that the value of a given language's nPVI can differ widely from one study to the next, likely due to differences in corpora, speech

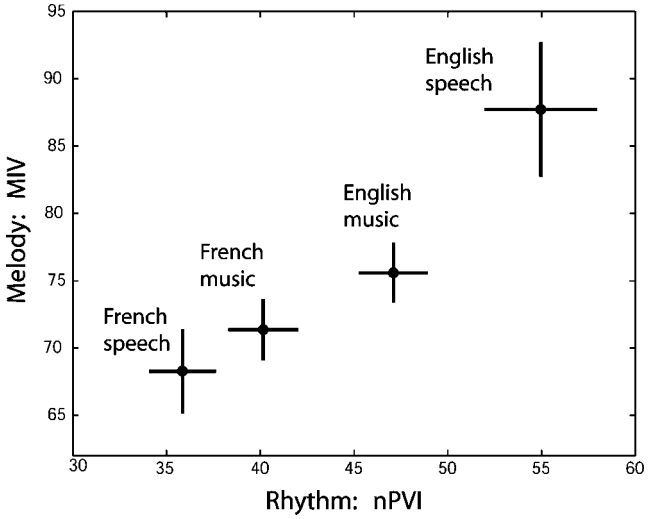

FIG. 6. Rhythm-melody (RM) space for speech and music. Axes are nPVI and MIV. Error bars show standard errors.

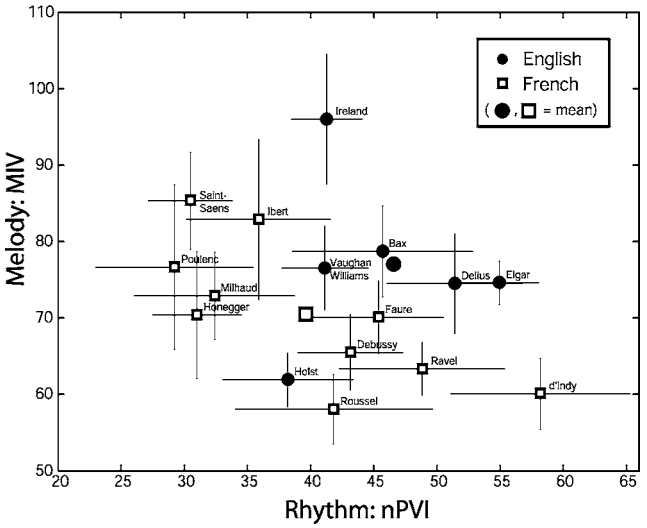

FIG. 7. nPVI and MIV values for individual composers. Error bars show standard errors. Note the almost complete separation of English and French composers in RM space, despite large overlap between the nationalities along either single dimension.

rate, and phonetic segmentation criteria. The same may be true of MIV values. Thus the position of a given language in RM space will likely vary from study to study. What is relevant is the *distance* between languages *within* a given study, where corpora and other criteria are more tightly controlled. The same point applies to music.

Using RM space, the "prosodic distance" (pd) between two languages can be defined as the Euclidean distance between the points representing the mean nPVI and MIV of the languages. For English and French speech (Es, Fs), this distance is

$$\mathrm{pd(Es,Fs)} = \sqrt{(\mathrm{nPVI_{Es}} - \mathrm{nPVI_{Fs}})^2 + (\mathrm{MIV_{Es}} - \mathrm{MIV_{Fs}})^2}. \quad (3)$$

In the current data the distance between English and French speech is 27.7 RM units. Applying the same equation to the musical data (i.e., replacing Es and Fs with Em and Fm), yields a distance between English and French music of 8.5 RM units. Thus the musical distance is about 30% of the linguistic difference.

Another aspect of Fig. 6 worth noting is that a line connecting English and French speech in RM space would lie at a very similar angle to a line connecting English and French music. In fact, if one defines vectors between the two languages and the two musics, using standard trigonometric formulas one can show that the angle between these vectors is only 14.2°. Thus a move from French to English speech in RM space involves going in a very similar direction as a move from French to English music.

Focusing now on the musical data, it is interesting to examine the position of individual composers in RM space, as shown in Fig. 7. Figure 7 reveals that English and French composers occupy distinct regions of RM space, despite large variation along any single dimension (Holst is the one exception, and is discussed later). This suggests that the *joint* properties of melody and rhythm, not either one alone, are involved in defining national characteristics of music.

IV. DISCUSSION

A. Aspects of speech prosody reflected in music

New tools from phonetics permit the confirmation of an old intuition shared by musicologists and linguists, namely that the instrumental music of a culture can reflect the prosody of its native language. These tools (the nPVI and the prosogram) are noteworthy because they allow quantitative comparisons of rhythm and melody in speech and music.

An exploration of the relationship between durational contrastiveness (nPVI) and durational variability reveals that nPVI differences between English and French speech are not a by-product of variability differences, even though the two languages do have significant differences in the variability of vowel durations in sentences. Variability differences also cannot account for musical nPVI differences between English and French (in fact, the two musics do not show a significant variability difference). It is thus clear that music reflects durational contrastiveness in speech, not durational variability. This finding is interesting in light of claims by linguists that part of the characteristic rhythm of English is a tendency for full and reduced vowels to alternate in spoken sentences (e.g., Bolinger, 1981). It appears that this tendency (or its absence) may be reflected in music.

Turning to melody, measures of pitch variability reveal that a specific aspect of speech melody is reflected in music, namely the variability of interval size between successive pitches in an utterance. English speech shows greater interval variability than French speech, even though the average pitch interval size in the two languages is nearly identical. This same pattern is reflected in music. Initially it may seem odd that pitch *intervals* in speech are reflected in music. While the human perceptual system is quite sensitive to interval patterns in music (where for example a melody can be recognized in transposition as long as its interval pattern is preserved), music features well-defined interval categories such as the minor second and perfect fifth while speech does not. Might the perceptual system attend to interval patterns in speech despite the lack of stable interval structure? Recent theoretical and empirical work in intonational phonology suggests that spoken pitch intervals may in fact be important in the mental representation of intonation, even if they do not adhere to fixed frequency ratios (Dilley, 2005). If this is the case, then one can understand why the perceptual system might attend to interval patterns in speech as part of learning the native language. It remains to be explained, though, why English speech should have a greater degree of interval variability than French. One idea is that British English may have three phonologically distinct pitch levels in its intonation system, while French may only have two (cf. Willems, 1982; Ladd and Morton, 1997; Jun and Fougeron, 2000; Jun, 2005). A compelling explanation, however, awaits future research.

While the current study focuses on just two cultures, it is worth noting that the techniques presented here are quite general. They can be applied to any culture where a sufficient sample of spoken and musical patterns can be collected. It would be quite interesting, for example, to use them to study relations between speech and instrumental music in

tone languages. Thai and Mandarin may be a good choice for such a comparison, since the languages have very different nPVI values (Grabe and Low, 2002) and different tone patterns, and since both languages come from cultures with well-developed instrumental music traditions.

B. A possible mechanism

By what route might speech patterns find their way into music? One oft-heard proposal is that composers borrow tunes from folk music, and that these tunes bear the stamp of linguistic prosody because they were written with words. This might be termed the "indirect route" from speech to music. The current study proposes a different hypothesis based on the idea of a "direct route" between the two domains. One advantage of the direct-route hypothesis is that it can account for the reflection of speech in music not thought to be particularly influenced by folk music (e.g., much of Elgar's and Debussy's work, cf. Grout and Palisca, 2000). The direct-route hypothesis centers on the notion of statistical learning of prosodic patterns in the native language. Statistical learning refers to tracking patterns in the environment and acquiring implicit knowledge of their statistical properties, without any direct feedback. Statistical learning of prosodic patterns in one's native language likely begins early in development. Research on language development has shown that infants are adept at statistical learning of phonetic/syllabic patterns in speech and of pitch patterns in nonlinguistic tone sequences (Saffran *et al.*, 1996; 1999) Thus it seems plausible that statistical learning of rhythmic and tonal patterns in speech would also begin in infancy, especially since infants are known to be quite sensitive to the prosodic patterns of language (Jusczyk, 1997; Nazzi *et al.*, 1998; Ramus 2002b).

Statistical learning of tone patterns need not be confined to infancy. Adult listeners show sensitivity to the distribution of different pitches and to interval patterns (Oram and Cuddy, 1995; Saffran *et al.*, 1999; Krumhansl, 2000; Krumhansl *et al.*, 2000). Importantly, statistical learning in music can occur with atonal or culturally unfamiliar materials, meaning that it is not confined to tone patterns that follow familiar musical conventions. Synthesizing these findings with the comments in the preceding paragraph leads to the hypothesis that statistical learning of the prosodic patterns of speech creates implicit knowledge of rhythmic and melodic patterns in language, which can in turn influence the creation of rhythmic and tonal patterns in music. Importantly, there is no claim that this influence is obligatory. Rather, linguistic patterns are seen as one resource available to a composer (either consciously or subconsciously) when they set out to compose music that is "of their culture" (cf. Patel and Daniele, 2003). Future work should consider what sort of evidence would support a causal link of this kind between speech and music.

C. Rhythm-melody relations: A domain for future investigation

Figure 7 suggested that joint properties of melody and rhythm may be particularly important for distinguishing between the music of different nations. While this study has examined rhythm and melody independently, it would be worth examining relations between rhythmic and melodic patterns in future work. Such relations may distinguish between languages and be reflected in music. One motivation for pursuing this issue is research in music perception suggesting that listeners are sensitive to such relations, e.g., the temporal alignment (or misalignment) of peaks in pitch and in duration (Jones, 1987; 1993). A second motivation is research on prosody which has revealed language-specific relations in the form of stable patterns of alignment of pitch peaks and valleys relative to the segmental string (Arvaniti *et al.*, 1998; Ladd *et al.*, 1999; Atterer and Ladd, 2004). This suggests that part of the characteristic "music" of a language is the temporal alignment between rhythmic and melodic patterns.

The key practical issue for future studies of rhythm-melody relations is how to represent pitch and temporal patterns in a manner that facilitates cross-domain comparison. Figure 8 illustrates one idea based on a vowel-based analysis of pitch and timing in speech. Panel (a) shows the pitch pattern of a French sentence ("Les mères sortent de plus en plus rapidement de la maternité") as a sequence of vowel

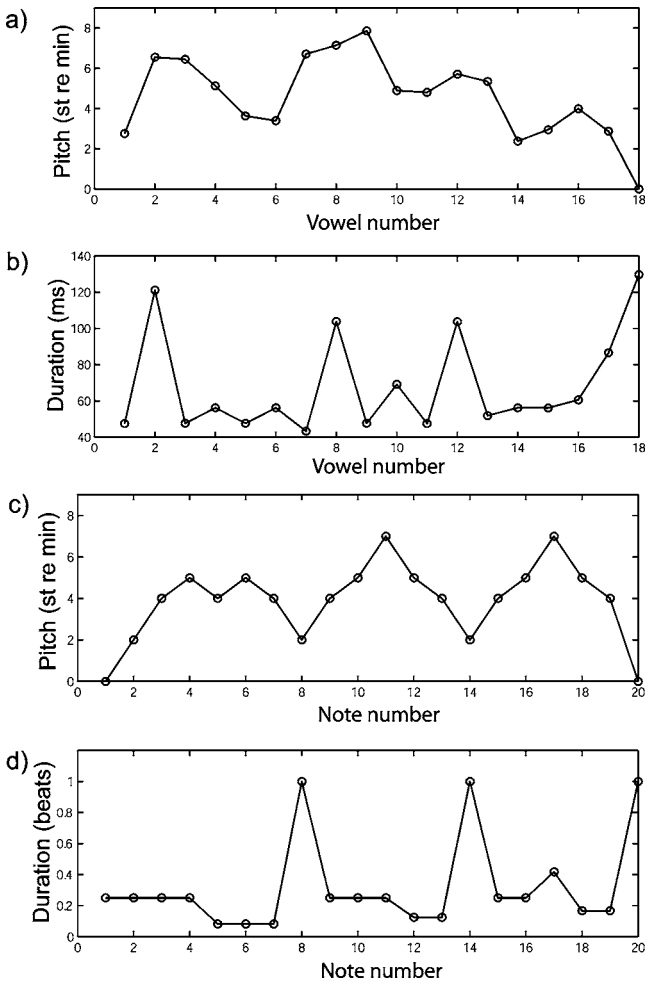

FIG. 8. (a, b) Pitch and duration patterns for the French sentence "Les mères sortent de plus en plus rapidement de la maternité" as uttered by a female speaker. (c, d) Pitch and duration patterns for a French musical theme from Debussy's *Les Parfums de la Nuit*. See text for details.

pitches, with each number representing the semitone distance of each vowel from the lowest vowel pitch in the sequence. Panel (b) shows the corresponding sequence of vowel durations (in ms). Panel (c) shows the pitch pattern of a French musical theme (4th theme of the 2nd movement of *Les Parfums de la Nuit*, theme D57 in Barlow and Morgenstern, 1983), using the same convention as panel (a), while panel (d) shows the duration pattern of the musical theme in fractions of a beat. The pitches in panel (a) are taken from the prosogram of the French sentence, but a reasonable estimate of these pitches can be made by simply taking the median Fo of each vowel (cf. Sec. IV D below).

When viewed in this way, it is easy to see how equivalent measurements of rhythm-melody relations can be made in speech and music. For example, one can ask whether there is a different pattern of alignment between pitch and duration peaks in language A than in language B, and if this difference is reflected in music. (This particular example is of interest to comparisons of British English and French, since phonologists have long noted that French sentences tend to be organized into phrases with pitch peaks and durational lengthening on the last syllable of each nonfinal phrase, e.g., Delattre, 1963; Jun and Fougeron, 2000; Di Cristo, 1998). Of course, one need not always look from speech to music; the direction of analysis can be reversed. In this case, one would first examine music for culturally-distinctive rhythm-melody relations, and then ask if these relations are reflected in speech. For example, the British composer Gustav Holst is unusual in being located in the "French" region of RM space in Fig. 6. Yet intuitively there is nothing French sounding about Holst's music (Indeed, the tune of "I Vow to Thee my Country," based on a theme from the *Jupiter* movement of *The Planets*, has often been suggested as an alternative national anthem for Britain). While Holst's anomalous position in RM space could simply be a sampling issue (i.e., it could change if more themes were added) it could also be that some as-yet unidentified relationship between rhythm and melody identifies his music as distinctively English. If so, it would be quite interesting to use this relationship to define a third dimension in RM space to see if it also separates English and French speech.

D. A possible application of RM space to quantifying non-native prosody

Learning to speak a second language with fluency requires acquisition of both segmental and prosodic patterns. Difficulty with L2 prosody is recognized as a challenge for language learners, particularly when L1 and L2 are rhythmically or melodically quite distinct (Pike, 1945; Chela-Flores, 1994). Yet there are very few quantitative methods for assessing non-native prosody. Such methods could prove useful for providing quantitative feedback to language learners in computer-based accent training programs (i.e., when practicing without the benefit of feedback from a human teacher), as well as for basic research on prosodic acquisition. RM space may have some useful qualities in this regard. As discussed in Sec. III C, it is possible to compute the prosodic distance between any two points in RM space [see Fig. 6 and Eq. (3)]. This means RM space could be used to quantify a speaker's prosodic distance from a target language which he or she is trying to acquire. For example, imagine that French and English represent $L1$ and $L2$ for a hypothetical language learner X. Further imagine that X is asked to read a set of sentences in $L2$, and nPVI and MIV measurements are made of these sentences based on the techniques used in the current study. X's mean nPVI and MIV values would define a point in RM space, whose prosodic distance from native English values, pd(X,Es), could be quantified using Eq. (3). This number could serve as the basis for a quantitative prosodic score indicating how close a given speaker was to native $L2$ prosody.

Should such an approach be taken, it will be important to determine how prosodic distance relates to perceptual judgments of foreign accent. For example, there may be a nonlinear relationship between these quantities. Only empirical research can resolve this issue. Fortunately, research relating quantitative measures of prosody to judgments of non-native accent has recently begun, motivated by the desire to test the perceptual relevance of different proposed rhythm measures. For example, White and Mattys (2005b) have studied the rhythm of non-native English as spoken by native Dutch and Spanish speakers. They examined a number of rhythm measures (such as ΔC, $\%V$, nPVI, etc.) and found that the best predictor of native-accent rating was a measure based on the coefficient of variation of vowel duration in sentences This measure, "VarcoV" (cf. Dellwo, 2006) showed an inverse linear relationship with degree of perceived foreign accent.[5] The methods of White and Mattys could be adapted to study the relationship between foreign accent judgments and measures of prosodic distance in RM space.

One notable feature of RM space is that it can easily be generalized to higher dimensions. Since RM space represents duration and pitch, an obvious candidate for an additional dimension is amplitude. In fact, a vowel-based amplitude measure has been shown to differentiate English and French speech. As part of a study comparing the variability of syllabic prominence in the two languages, Lee and Todd (2004) measured the intensity (rms) variation among vowels in a sentence, computing a value ΔI for each sentence, representing the standard deviation of the intensity values. ΔI was significantly larger for English than for French. Inspired by this work, the current study measured the RMS of each vowel in the current corpus (after first normalizing each sentence to 1 V rms). The CV of vowel rms was then computed within each sentence. Consistent with Lee and Todd's (2004) findings, English had a significantly higher vowel rms variability than French (mean and s.e. were 0.53 (0.02) and 0.40 (0.02) for English and French respectively, $p<0.001$). Thus English and French are distinct in a three-dimensional RM space with nPVI, MIV, and rms variation as orthogonal axes. Prosodic distances can be computed in $3d$ RM space via a generalization of Eq. (3) to three dimensions, and can be used to measure a non-native speaker's prosodic distance from a target language in three dimensions.

Whether using two or three dimensions, the major practical challenge in using RM space is phonetic segmentation of sentences, a time-consuming endeavor when done by hu-

mans. This problem can be alleviated by having speakers read a fixed set of sentences whose texts are known. In this case vowel boundaries can be marked in speech signals using the procedure of forced alignment, in which a speech recognizer is given the list of segments in the order in which they appear in the signal. Dellwo (personal communication) has used this technique with the HTK speech recognition software and found very high agreement with human labeling for speech at a normal rate.

Once vowel boundaries are placed within utterances, then nPVI and MIV measurements can be made based on existing algorithms. It is even possible to compute an accurate estimate of MIV without computing prosograms, by simply assigning each vowel a level pitch based on its median Fo and computing intervals from these pitches. (To illustrate the accuracy of this approach, the resulting mean and s.e. of MIV for English and French in the current corpus were 87 (4) and 68 (3), respectively, compared to 88 (5) and 68 (3) for prosogram-based measures, cf. Table III.) In this case, all that is needed to create RM-space plots such as Fig. 6 are sentences with vowel boundaries marked and Fo contours extracted, together with software that can compute the median Fo of each vowel based on this information. The remaining nPVI and MIV computations can be done via simple equations in a spreadsheet.

V. CONCLUSION

The rhythms and melodies of speech and instrumental music can be quantitatively compared using tools from modern phonetics. Using these tools, an investigation of language and music from England and France confirms the intuition that music reflects the prosody of a composer's native language. The approaches developed here can be applied to the study of language-music relations in other cultures, and may prove useful in quantifying non-native prosody.

ACKNOWLEDGMENTS

We thank Franck Ramus for providing the English and French sentences (audio files and phonetic transcriptions), Piet Mertens for help with prosograms, Peter Ladefoged for help with IPA transcription of the English sentences, Laura Dilley and D. Robert Ladd for helpful comments on this manuscript, and Philip Ball for drawing our attention to "I vow to thee my country." This work was supported by the Neurosciences Research Foundation as part of its research program on music and the brain at The Neurosciences Institute, where A.D.P. is the Esther J. Burnham Fellow and J.R.I. is the Karp Foundation Fellow, and by a grant from the H.A. and Mary K. Chapman Charitable Trust.

APPENDIX A: ENGLISH AND FRENCH SENTENCES

A hurricane was announced this afternoon on the TV.

My grandparent's neighbor's the most charming person I know.

Much more money will be needed to make this project succeed.

The local train left the station more than 5 minutes ago.

The committee will meet this afternoon for a special debate.

The parents quietly crossed the dark room and approached the boy's bed.

This supermarket had to close due to economic problems.

In this famous coffee shop you will eat the best donuts in town.

This rugby season promises to be a very exciting one.

Science has acquired an important place in western society.

The last concert given at the opera was a tremendous success.

In this case, the easiest solution seems to appeal to the court.

Having a big car is not something I would recommend in this city.

They didn't hear the good news until last week on their visit to their friends.

Finding a job is difficult in the present economic climate.

The library is open every day from 8 a.m. to 6 p.m.

The government is planning a reform of the education program.

This year's Chinese delegation was not nearly as impressive as last year's.

The city council has decided to renovate the Medieval center.

No welcome speech will be delivered without the press offices' agreement.

Les parents se sont approchés de l'enfant sans faire de bruit.

Cette boulangerie fabrique les meilleurs gâteaux de tout le quartier.

La femme du pharmacien va bientôt sortir faire son marché.

Les voisins de mes grandparents sont des personnes très agréables.

Il faudra beaucoup plus d'argent pour mener à bien ce projet.

Le magasin est ouvert sans interruption toute la journée.

Les mères sortent de plus en plus rapidement de la maternité.

L'été sera idyllique sur la côte méditerranéenne.

Ils ont appris l'évènement au journal télévisé de huit heures.

La nouvelle saison théâtrale promet d'être des plus intéressante.

Un tableau de très grande valeur a été récemment dérobé.

Le plus rapide est encore le recours auprès de la direction.

Les récents événements ont bouleversé l'opinion internationale.

Le train express est arrivé en gare il y a maintenant plus de 5 minutes.

La reconstruction de la ville a commencé après la mort du roi.

L'alcool est toujours la cause d'un nombre important d'accidents de la route.

Aucune dérogation ne pourra être obtenue sans l'avis du conseil.

Les banques ferment particulièrement tôt le vendredi soir.

Trouver un emploi est difficile dans le contexte économique actuel.

Le ministère de la culture a augmenté le nombre de ces subventions.

APPENDIX B: COMPOSERS AND MUSICAL THEMES

Code numbers are those used in Barlow and Morgenstern (1983).

English: **Bax** b508, b509, b510, b511, b515, b517, b518, b519, b520. **Delius** d189, d191, d192, d193, d194, d195, d196, d197, d198, d199, d200, d201, d202, d205, d208, d214, d215, d216, d219. **Elgar** e3, e4, e7, e8, e13, e14, e15, e16, e17, e18, e19, e20, e21, e23, e27, e28, e30, e31, e33, e34, e35, e51, e52, e53, e56, e58, e60, e61, e62, e63, e64, e66, e67, e68, e70, e71, e72, e73a, e73b, e73c, e73d, e73f, e73h, e73i, e73j. **Holst** h798, h799, h801, h803, h804, h805,

h806, h807, h810, h811, h813, h814, h817, h818, h819, h820. **Ireland** i95, i97, i98, i102, i104, i105, i109, i110, i111, i112, i113. **Vaughan Williams** v4, v5, v6, v7, v8, v12, v13, v14, v17, v18, v19, v20, v21, v22, v23, v24, v26, v27, v28, v29, v30, v31, v32, v33, v34, v35, v37, v38, v39, v40, v41, v42, v43, v44, v45, v49.

French: **Debussy** d13, d14, d20, d21, d42, d43, d55, d57, d58, d62, d70, d71, d74, d77, d78, d80, d83, d85, d86, d87, d88, d90, d97, d98, d100, d105, d107, d108, d109, d113, d116, d117, d118, d122, d123, d124, d125, d126, d127, d129, d132, d134, d135, d138, d139, d140. **Fauré** f60, f61, f62, f63, f72, f75, f76, f76d, f77, f78, f79, f80, f84, f85, f87, f89, f91, f92, f93, f94, f95, f97, f98, f101, f102, f103, f104, f105. **Honegger** h830, h832, h833, h834, h836, h842, h843, h844. **Ibert** i1, i3, i4, i6, i8, i13, i14, i24, i26, i27. **D'Indy** i31, i33, i40, i41, i42, i44, i47, i48. **Milhaud** m382, m383, m384, m386, m387, m394, m395. **Poulenc** p170, p171, p176, p177, p178. **Ravel** r124, r128, r129, r130, r132, r133, r147, r148, r150, r151, r152, r153, r154, r155, r156, r183, r184, r186. **Roussel** r407, r409, r410, r411, r412, r416, r417, r419, r420, r422, r423. **Saint-Saëns** s18, s20, s21, s22, s26, s31, s32, s33, s34, s35, s36, s40, s42, s49, s50, s66, s69, s77, s79, s89, s92, s98, s99, s100, s102, s103, s104, s105, s106, s107, s108, s109, s110, s112, s114, s127, s129, s133, s134.

[1]In the case of Thai and other languages with phonemic vowel length contrast, a high nPVI value may be driven by these length contrasts in addition to (or rather than) vowel reduction.

[2]It should be noted that all published nPVI studies comparing English and French have focused on speakers of the standard dialect of each language. Research on dialectal variation in English speech rhythm suggests that within-language variation in nPVI is smaller than the nPVI difference between English and French (E. Ferragne, unpublished data; White and Mattys, 2005a), but dialectical rhythm studies in both languages are needed to establish whether within-language variation is smaller than between-language variation.

[3]As with the better-known autosegmental-metrical approach, the prosogram is an abstraction of an Fo curve. The current study prefers the prosogram to the autosegmental-metrical abstraction because it is explicitly concerned with the psychoacoustics of intonation perception, and thus seems better suited to comparing speech and music as patterns of perceived pitches. That being said, it should be noted that the prosogram is based on research on native listeners of intonation languages, and hence its applicability to pitch patterns in tone languages is uncertain.

[4]The prosogram is freely available from http://bach.arts.kuleuven.be/pmertens/prosogram/, and runs under Praat, which is freely available from http://www.fon.hum.uva.nl/praat/

[5]It should be noted that VarcoV was only slightly better than nPVI as a predictor of accent ratings in this study. Furthermore, a concern about this research is that segmental and suprasegmental cues are probably both contributing to foreign accent ratings. Thus some method is needed to remove interspeaker variability in segmental cues, such as speech resynthesis. In fact, White and Mattys (personal communication) are pursuing such an approach.

Abraham, G. (**1974**). *The Tradition of Western Music* (University of California Press, Berkeley), Chap. 4, pp. 61–83.

Arvaniti, A., Ladd, D. R., and Mennen, I. (**1998**). "Stability of tonal alignment: The case of Greek prenuclear accents," J. Phonetics **26**, 3–25.

Atterer, M. and Ladd, D. R. (**2004**). "On the phonetics and phonology of 'segmental anchoring' of Fo: Evidence from German," J. Phonetics **32**, 177–197.

Barlow, H. and Morgenstern, S. (**1983**). *A Dictionary of Musical Themes*, revised ed. (Faber and Faber, London).

Beckman, M. (**1992**). "Evidence for speech rhythm across languages," in *Speech Perception, Production, and Linguistic Structure*, edited by Y. Tohkura *et al.* (IOS, Tokyo), pp. 457–463.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in Proceedings of the Institute Of Phonetic Sciences, University of Amsterdam, Vol. **17**, pp. 97–110.

Bolinger, D. (**1981**). *Two Kinds of Vowels, Two Kinds of Rhythm* (Indiana University Linguistics Club, Bloomington).

Chela-Flores, B. (**1994**). "On the acquisition of English rhythm: Theoretical and practical issues," IRAL **32**, 232–242.

d'Alessandro, C. and Castellengo, M. (**1994**). "The pitch of short-duration vibrato tones," J. Acoust. Soc. Am. **95**, 1617–1630.

d'Alessandro, C. and Mertens, P. (**1995**) "Automatic pitch contour stylization using a model of tonal perception," Comput. Speech Lang. **9**, 257–288.

Daniele, J. R. and Patel, A. D. (**2004**). "The interplay of linguistic and historical influences on musical rhythm in different cultures," in Proceedings of the 8th International Conference on Music Perception and Cognition, Evanston, IL, pp. 759–762.

Dauer, R. M. (**1983**). "Stress-timing and syllable-timing reanalyzed," J. Phonetics **11**, 51–62.

Dauer, R. M. (**1987**). "Phonetic and phonological components of language rhythm," in Proceedings of the 11th International Congress of Phonetic Sciences, Tallinn, Vol. **5**, pp. 447–450.

Delattre, P. (**1963**). "Comparing the prosodic features of English, German, Spanish and French," IRAL **1**, 193–210.

Delattre, P. (**1966**). "A comparison of syllable length conditioning among languages," IRAL **4**, 183–198.

Delattre, P. (**1969**). "An acoustic and articulatory study of vowel reduction in four languages," IRAL **7**, 295–325.

Dellwo, V. (**2006**). "Rhythm and Speech Rate: A variation coefficient for deltaC," in *Language and Language-processing. Proceedings of the 38th linguistic Colloquium, Piliscsaba 2003*, edited by P. Karnowski and I. Szigeti(Peter Lang, Frankfurt am Main), pp. 231–241

Dellwo, V., Steiner, I., Aschenberner, B., Dankovičová, J., and Wagner, P. (**2004**). "The BonnTempo-Corpus and BonnTempo-Tools: A database for the study of speech rhythm and rate," in Proceedings of the 8th ICSLP, Jeju Island, Korea.

Di Cristo, A. (**1998**). "Intonation in French," in *Intonation Systems: A Survey of Twenty Languages*, edited by D. Hirst and A. Di Cristo (Cambridge University Press, Cambridge), pp. 195–218.

Dilley, L. (**2005**). "The phonetics and phonology of tonal systems," Ph.D. dissertation, MIT.

Grabe, E., Post, B., and Watson, I. (**1999**). "The acquisition of rhythmic patterns in English and French," in Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, pp. 1201–1204.

Grabe, E. and Low, E. L. (**2002**). "Durational variability in speech and the rhythm class hypothesis," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 515–546.

Grout, D. J. and Palisca, C. V. (**2000**). *A History of Western Music* 6th ed. (Norton, New York).

Hall, R. A., Jr. (**1953**). "Elgar and the intonation of British English," The Gramophone, June 1953:6-7, reprinted in *Intonation: Selected Readings*, edited by D. Bolinger (Penguin, Harmondsworth), pp. 282–285.

Hermes, D. and van Gestel, J. C. (**1991**). "The frequency scale of speech intonation," J. Acoust. Soc. Am. **90**, 97–102.

House, D. (**1990**) *Tonal Perception in Speech* (Lund University Press, Lund).

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

Jones, M. R. (**1987**). "Dynamic pattern structure in music: recent theory and research," Percept. Psychophys. **41**, 621–634.

Jones, M. R. (**1993**). "Dynamics of musical patterns: How do melody and rhythm fit together?" in *Psychology and Music: The Understanding of Melody and Rhythm*, edited by T. J. Tighe and W. J. Dowling (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 67–92.

Jusczyk, P. (**1997**). *The Discovery of Spoken Language* (MIT Press, Cambridge, MA).

Jun, S.-A. (**2005**). "Prosodic Typology," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, edited by S-A. Jun (Oxford University Press, Oxford), pp. 430–458.

Jun, S-A. and Fougeron, C. (**2000**) "A Phonological Model of French Intonation," in *Intonation: Analysis, Modeling and Technology*, edited by A. Botinis (Kluwer Academic, Dordrecht), pp. 209–242.

Krumhansl, C. (**2000**). "Tonality induction: A statistical approach applied

cross-culturally," Music Percept. **17**, 461–479.
Krumhansl, C., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., and Louhivuori, J. (**2000**). "Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks," Cognition **76**, 13–58.
Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge).
Ladd, D. R. and Morton, R. (**1997**). "The perception of intonational emphasis: Continuous or categorical?," J. Phonetics **25**, 313–342.
Ladd, D. R., Faulkner, D., Faulkner, H., and Schepman, A. (**1999**). "Constant 'segmental anchoring' of Fo movements under changes in speech rate," J. Acoust. Soc. Am. **106**, 1543–1554.
Lee, C. S. and Todd, N. P. McA. (**2004**). "Toward an auditory account of speech rhythm: Application of a model of the auditory 'primal sketch' to two multi-language corpora," Cognition **93**, 225–254.
Lerdahl, F. and Jackendoff, R. (**1983**). *A Generative Theory of Tonal Music* (MIT Press, Cambridge, MA).
Liberman, M. (**1975**). "The intonational system of English," Ph.D. thesis, MIT.
Low, E. L. (**1998**). "Prosodic prominence in Singapore English," Ph.D. thesis, University of Cambridge.
Low, E. L., Grabe, E., and Nolan, F. (**2000**). "Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English," Lang Speech **43**, 377–401.
Maidment, J. (**1976**). "Voice fundamental frequency characteristics as language differentiators," Speech and Hearing, Work in Progress, Univ. College London, pp. 75–93.
Mehler, J., Dupuox, E., Nazzi, T., and Dehaene-Lambertz, D. (**1996**). "Coping with linguistic diversity: The infant's viewpoint," in *Signal to Syntax*, edited by J. L. Morgan and D. Demuth (Lawrence Erlbaum, Mahwah, NJ), pp. 101–116.
Mertens, P. (**2004a**). "The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model," in Proceedings of Speech Prosody 2004, Nara, Japan, pp. 23–26.
Mertens, P. (**2004b**). "Un outil pour la transcription de la prosodie dans les corpus oraux," Traitement Automatique des Langues **45**, 109–130.
Mertens, P. and d'Alessandro, C. (**1995**). "Pitch contour stylization using a tonal perception model," in Proceedings of the 13th International Congress on Phonetic Sciences, Stockholm, pp. 228–231.
Mertens, P., Beaugendre, F., and d'Alessandro, C. (**1997**). "Comparing approaches to pitch contour stylization for speech synthesis," in *Progress in Speech Synthesis*, edited by J. P. H. van Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg (Springer Verlag, New York), pp. 347–363.
Nazzi, T., Bertoncini, J., and Mehler, J. (**1998**). "Language discrimination in newborns: Toward an understanding of the role of rhythm," J. Exp. Psychol. Hum. Percept. Perform. **24**, 756–777.
Nespor, M. (**1990**). "On the rhythm parameter in phonology." in *Logical Issues in Language Acquisition*, edited by I. Rocca (Foris, Dordrecht), pp. 157–175.
Nolan, F. (**2003**). "Intonational equivalence: An experimental evaluation of pitch scales," in Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, pp. 771–774.
Oram, N. and Cuddy, L. L. (**1995**). "Responsiveness of Western adults to pitch distributional information in melodic sequences," Psychol. Res. **57**, 103–118.
Palmer, C. (**1997**). "Music performance," Annu. Rev. Psychol. **48**, 115–138.
Patel, A. D. and Daniele, J. R. (**2003**). "An empirical comparison of rhythm in language and music," Cognition **87**, B35–B45.
Patel, A. D. and Daniele, J. R. (**2003b**). "Stress-timed vs syllable-timed music? A comment on Huron and Ollen," Music Percept. **21**, 273–276.
Peterson, G. E. and Lehiste, I. (**1960**). "Duration of Syllabic Nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.
Pierrehumbert, J. (**1980**). "The phonetics and phonology of English intonation," Ph.D. dissertation, MIT.
Pike, K. N. (**1945**). *The Intonation of American English* (University of Michigan, Ann Arbor).
Ramus, F., Nespor, M., and Mehler, J. (**1999**). "Correlates of linguistic rhythm in the speech signal," Cognition **73**, 265–292.
Ramus, F. (**2002**). "Acoustic correlates of linguistic rhythm: Perspectives," in Proceedings of Speech Prosody, Aix-en-Provence, pp. 115–120.
Ramus, F. (**2002b**). "Language discrimination by newborns: teasing apart phonotactic, rhythmic, and intonational cues," Annual Review of Language Acquisition **2**, 85–115.
Repp, B. H. (**1992**). "Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's 'Träumerei'," J. Acoust. Soc. Am. **92**, 2546–2568.
Roach, P. (**1982**). "On the distinction between 'stress-timed' and 'syllable-timed' languages," in *Linguistic Controversies: Essays in Linguistic Theory and Practice in Honour of F.R. Palmer*, edited by D. Crystal (Edward Arnold, London), pp. 73–79.
Rossi, M. (**1971**). "Le seuil de glissando ou seuil de perception des variations tonales pour la parole," Phonetica **23**, 1–33.
Rossi, M. (**1978a**). "La perception des glissandos descendants dans les contours prosodiques," Phonetica **35**, 11–40.
Rossi, M. (**1978b**). "Interactions of intensity glides and frequency glissandos," Lang Speech **21**, 384–396.
Sadakata, M. and Desain, P., "Comparing rhythmic structure in popular music and speech" (submitted).
Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996**). "Statistical learning by 8-month old infants," Science **274**, 1926–1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (**1999**). "Statistical learning of tone sequences by human infants and adults," Cognition **70**, 27–52.
Selkirk, E. O. (**1984**). *Phonology and Syntax: The Relation Between Sound and Structure* (MIT Press, Cambridge, MA).
'tHart, J. (**1976**). "Psychoacoustic backgrounds of pitch contour stylization." I. P. O. Annual Progress Report **11**, 11–19.
'tHart, J., Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody* (Cambridge University Press, Cambridge).
Wenk, B. J. (**1987**). "Just in time: On speech rhythms in music," Linguistics **25**, 969–981.
White, L. S. and Mattys, S. L. (**2005a**). "Calibrating rhythm: A phonetic study of British Dialects." in Proceedings of the Vth Conference on UK Language Variation and Change, Aberdeen, Scotland.
White, L. S. and Mattys, S. L. (**2005b**). "How far does first language rhythm influence second language rhythm?," in Proceedings of Phonetics and Phonology in Iberia, Barcelona, Spain.
Willems, N. (**1982**). *English Intonation from a Dutch Point of View* (Foris, Dordrecht).