



Brief article

An empirical comparison of rhythm in language and music

Aniruddh D. Patel*, Joseph R. Daniele

The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121, USA

Received 3 September 2002; accepted 1 October 2002

Abstract

Musicologists and linguists have often suggested that the prosody of a culture's spoken language can influence the structure of its instrumental music. However, empirical data supporting this idea have been lacking. This has been partly due to the difficulty of developing and applying comparable quantitative measures to melody and rhythm in speech and music. This study uses a recently-developed measure for the study of speech rhythm to compare rhythmic patterns in English and French language and classical music. We find that English and French musical themes are significantly different in this measure of rhythm, which also differentiates the rhythm of spoken English and French. Thus, there is an empirical basis for the claim that spoken prosody leaves an imprint on the music of a culture. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Speech rhythm; Musical rhythm; Stress-timing; Syllable-timing

1. Introduction

Speech and music are universal among human cultures. Both involve organized acoustic sequences and engage complex cognitive and motor processes. The relationship between language and music has long interested scholars across a broad range of disciplines, from linguistics to neuroscience, and is the focus of an expanding body of empirical research (e.g. Besson, Faiïta, Peretz, Bonnel, & Requin, 1998; Palmer & Kelly, 1992; Patel, Gibson, Ratner, Besson, & Holcomb, 1998a; Patel, Peretz, Tramo, & Labrecque, 1998b). Two main loci of interest are prosodic and syntactic structure. Comparisons of prosodic structure examine the way duration, pitch and intensity create structured rhythmic and melodic patterns in speech and music (e.g. Jusczyk & Krumhansl, 1993; Lerdahl & Jackendoff, 1983). Comparisons of syntactic structure examine the way discrete elements

* Corresponding author.

E-mail address: apatel@nsi.edu (A.D. Patel).

combine in hierarchical fashion to form coherent sequences in the two domains (Patel, 1998; Swain, 1997). The current study is concerned with a prosodic comparison of language and music.

A number of musicologists and linguists have claimed that the prosody of a composer's native language can influence the structure of his or her instrumental music (e.g. Abraham, 1974; Wenk, 1987). However, to date no satisfactory empirical evidence exists to support this claim. This lack of evidence reflects the fact that it has been difficult to meet all of the criteria necessary for testing this idea. First, a measure is needed to quantify prosodic structure in one or more languages. Second, this same measure needs to be applicable to music so that language and music can be compared in a common framework. Third, samples of speech and music must be broad enough to ensure that the phenomena are not idiosyncratic to particular speakers or composers. A primary goal of this study is to meet all of these requirements.

We pursue the hypothesis that music reflects linguistic prosody by focusing on speech and music from two countries with very different language rhythms: Britain and France. Linguistic "rhythm" refers to the way languages are organized in time. British English and standard French are widely acknowledged by linguists to have very different rhythmic organization, representing "stress-timed" and "syllable-timed" languages, respectively (Abercrombie, 1967; Pike, 1945). It has been hypothesized that stress-timed languages have equal duration between stresses, while syllable-timed languages show equal duration between syllable onsets. Empirical research fails to support this hypothesis, leading one researcher to lament that "one is obliged to conclude that the basis for the distinction [between linguistic rhythm classes] is auditory and subjective..." (Roach, 1982). However, recent phonetic work has demonstrated that there are indeed quantitative rhythmic differences between stress- and syllable-timed languages. This work has set aside the idea of isochrony and has focused instead on the durational patterning of vocalic and intervocalic intervals in speech (Grabe & Low, 2002; Low, Grabe, & Nolan, 2000; Ramus, Nespor, & Mehler, 1999).¹ One basic intuition behind this work is that stress-timed languages show a greater degree of vowel reduction than syllable-timed languages (Dauer, 1983, 1987; Nespor, 1990), suggesting that the variability of vowel duration should be greater in stress- vs. syllable-timed languages (Grabe & Low, 2002). Another idea is that stress-timed languages tend to permit complex syllables (e.g. the English word "strengths" is CCCVCCC), and thus may have greater durational variability in consonant sequences than syllable-timed languages (Ramus et al., 1999).

These ideas have been borne out in empirical research. Ramus et al. (1999) have shown that the durational variability of consonantal intervals is greater in stress-timed languages such as British English and Dutch than in syllable-timed languages such as Spanish, French, and Italian. Grabe and Low (2002) have recently found that vocalic duration is more variable in British English, Dutch, and German (another stress-timed language) than in French and Spanish. With regard to the latter finding, it is known that listeners are

¹ "Vocalic intervals are vowels and sequences of consecutive vowels, regardless of whether they belong to the same syllable (or word, for that matter) or not. Similarly, inter-vocalic or consonantal intervals are made up of consonants and sequences of consecutive consonants." (Ramus, 2002). For details of measurement criteria, see Ramus et al. (1999).

sensitive to the patterning of vowel duration in speech (i.e. this factor plays a role in judgments of the naturalness of synthetic speech; Carlson, Granström, & Klatt, 1979). Thus, the temporal patterning of vowels is likely to be one factor contributing to the rhythmic feel of a language.

The measure of durational variability used by Grabe and Low is particularly interesting with regard to speech–music comparisons. This is the “normalized Pairwise Variability Index” (nPVI), defined as:

$$\text{nPVI} = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right|$$

where m is the number of vocalic intervals in an utterance and d_k is the duration of the k th interval. Two aspects of this measure make it appealing for use with music. First, the nPVI is a purely relative measure of variability, i.e. the durational difference between each pair of intervals is measured *relative to* the average length of the pair. This normalization (which was originally introduced to control for fluctuations in speech rate) makes the nPVI a dimensionless quantity which can be applied to both language and music.² Second, the nPVI has been applied to vowels. Vowels form the core of syllables, which can in turn be compared to musical tones (i.e. in setting words to music it is quite common for each note to be assigned to one syllable). Our strategy, then, was to apply the nPVI to tone sequences from British and French instrumental music to determine if differences emerged which reflected the rhythmic differences between British English and French speech.

2. Methods

2.1. Background: rhythmic differences between British English and standard French speech

nPVI values for the variability of vocalic durations in English and French were obtained from a recent study by Ramus (2002) based on 20 utterances per language generated by four individuals per language speaking five sentences each (cf. Nazzi, Bertoncini, & Mehler, 1998; Ramus et al., 1999).³ The sentences consisted of short, news-like utterances read in a declarative tone of voice, and matched for number of syllables (15–19) and duration (about 3 s, see Table 1). The values obtained by Ramus are shown in Fig. 1. As can be seen from the figure, the nPVI for British English is greater than for French (means of 66.99 vs. 49.27, respectively). The difference is highly significant (Mann–Whitney U -test, $U = 66$, $P < 0.001$).

² Like the standard deviation, the nPVI measures variability in a set of observations. However, the nPVI focuses on variation between *successive* intervals, making it sensitive to the order of observations (cf. Low et al., 2000). This feature is unlikely to be crucial to the results of this study: other normalized measures of variability (e.g. the coefficient of variation) would likely lead to similar results.

³ Ramus' database is currently the largest published set of nPVI values for British English and French.

Table 1

Examples of sentences used to compute linguistic nPVI values (from Ramus, 2002)

English:

- The local train left the station more than five minutes ago.
- The last concert given at the opera was a tremendous success.
- The city council has decided to renovate the medieval center.

French:

- Les parents se sont approchés de l'enfant sans faire de bruit.
- Ils ont appris l'évènement au journal télévisé de huit heures.
- Le ministère de la culture a augmenté le nombre de ces subventions.

2.2. Choosing musical materials

Our source of musical material was a standard reference work in musicology, *A Dictionary of Musical Themes, 2nd Edition* (Barlow & Morgenstern, 1983), which focuses on the instrumental music of Western European composers. In choosing composers to include in our study, we were guided by two factors. First, the composers had to be from a relatively recent musical era, since measurements of speech prosody are based on contemporary speech, and languages are known to change over historical time in terms of sound structure. Second, the composers must have been native speakers of British English or French, who lived and worked in England or France. Using these guidelines, we examined all English and French composers from Barlow and Morgenstern who were born in the 1800s and died in the 1900s, and who had at least five musical themes in the dictionary which were eligible for inclusion in the study (see Section 2.3 for inclusion criteria). We chose composers who spanned the turn of the century because this is a time noted by musicologists as an important era for “musical nationalism”, i.e. the development of national

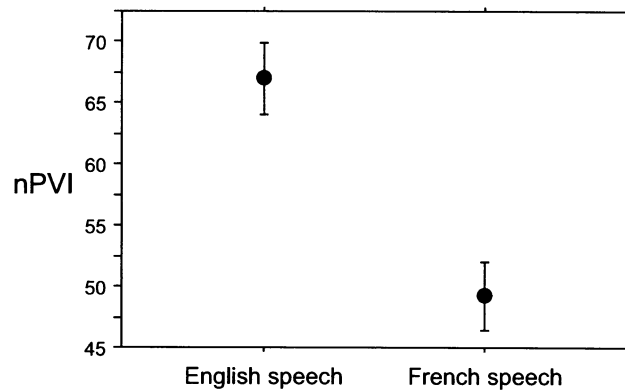


Fig. 1. Linguistic nPVI values for sentences in British English and standard French, from Ramus (2002) (English: mean = 66.99, SE = 2.89, $n = 20$; French: mean = 49.27, SE = 2.78, $n = 20$).

Table 2
Composers examined in this study

	Dates lived	# Themes	Example source
<i>English composers</i>			
Arnold Bax	1883–1953	9	Sonata, Viola & Piano
Frederick Delius	1862–1934	20	Concerto, Violin & Orchestra
Edward Elgar	1857–1934	41	Concerto in B minor
Gustav Holst	1874–1934	18	The Planets
John Ireland	1879–1962	11	Sonata in G Minor
Ralph Vaughan Williams	1872–1958	38	A London Symphony
<i>French composers</i>			
Claude Debussy	1862–1918	47	Les Parfums de La Nuit
Vincent D'Indy	1851–1931	12	Le Camp de Wallenstein
Gabriel Fauré	1845–1924	25	Quartet in C minor
Arthur Honegger	1892–1955	8	Pastorale D'Été, Orchestra
Jacques Ibert	1890–1962	12	Concerto, Alto Sax
Darius Milhaud	1892–1974	7	Pastorale for Ob., Cl., Bsn.
Francis Poulenc	1899–1963	5	Toccato, Piano
Maurice Ravel	1875–1937	18	Introd. & Allegro.
Albert Roussel	1869–1937	11	Symphony No. 3 in G minor
Camille Saint-Saëns	1835–1921	36	Concerto No. 3 in B minor

styles, in which speech prosody has been thought to play a role. Based on our criteria, six English and ten French composers were included in the study (Table 2).

2.3. Criteria for inclusion of musical themes

Since our analysis focused on instrumental music (i.e. music written for instruments, not for the voice), we excluded themes from pieces with “song” or similar terms in the title (i.e. barcarole, chanson, chant, choral, chorale, and serenade). Also, since we focused on rhythm, we were careful to exclude pieces whose titles indicated that they had an external rhythmic agenda (i.e. stylized dances such as waltzes, gigue, and gavottes, as well as marches). We also excluded pieces whose titles suggested that the composer was consciously striving for a special style (i.e. children’s music, “exotic” music of a foreign land or culture, or music styled on the work of another composer). Finally, we required that each theme have at least 12 notes and no internal pauses/rests (to provide a good sample for nPVI calculation), and not have any grace notes or fermatas (musical markings which introduce durational uncertainties). A total of 137 English musical themes and 181 French musical themes passed these criteria.

2.4. Measurement of musical nPVI values

nPVI values were computed directly from music notation (Fig. 2). To compute the nPVI for a given theme, the first note was arbitrarily given a duration of 1, and the durations of remaining notes were expressed as multiples or fractions of the first note. This string of values was then entered into the nPVI equation. Since the nPVI is a relative measure of

variability, the same nPVI value is obtained by any scheme which preserves the relative durations of notes.⁴

Initially it might seem objectionable to measure musical nPVI values from scores rather than from performed music, especially since linguistic nPVI values are (necessarily) based on acoustic measurements. However, measurement of actual musical performances raises a number of questions without simple answers. Most notably, which performance of the piece is to be measured, and how is this choice to be defended against all other recorded performances, each of which will differ in the precise timing of notes? This and other questions make it clear that the use of a composer's notation is actually a good choice for computing musical nPVI values, because the notation at least contains an unambiguous record of the composer's choice of relative note durations.

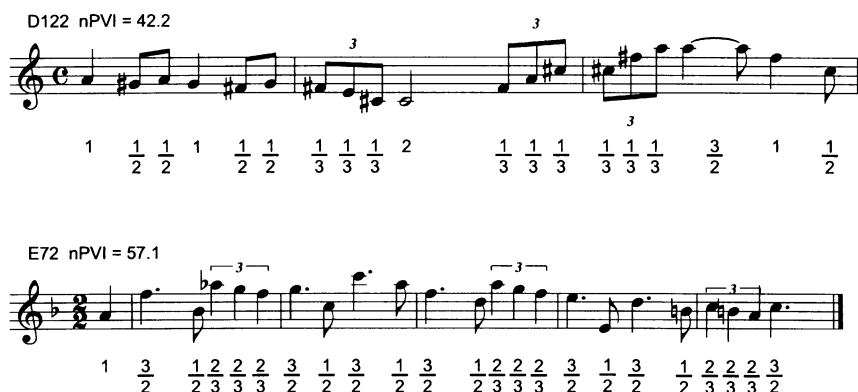


Fig. 2. An example of a Debussy theme (D122, *Quartet in G minor for Strings, 1st movement, 2nd theme*) and an Elgar theme (E72, *Symphony No. 1, in A Flat, Opus 55, 4th movement, 2nd theme*), together with the nPVI for each theme. The relative duration of each note is shown below the musical staff. The first note is assigned a duration of 1, and the durations of the remaining notes are expressed as multiples or fractions of this value.

3. Results

Fig. 3 shows the results of musical nPVI measurements in a format similar to the results of linguistic nPVI measurements (cf. Fig. 1). As can be seen, the average nPVI values for English and French musical themes are different, with English music having the greater value (mean nPVI = 46.91 vs. 40.90). This difference is significant (Mann–Whitney *U*-test, $U = 9993.5$, $P < 0.01$).

One important question about these results is the possible influence of musical meter on nPVI values. That is, if themes in binary vs. ternary meters (i.e. meters which divide beats into two vs. three subdivisions, such as 2/4 vs. 6/8; cf. London, 2002) tend to have different nPVI values, and the English and French themes in this study have different proportions of these metrical types, then this might account for the observed differences. To check this

⁴ For example, one could assign a quarter note a duration of 1 in all themes, and express other note durations as multiples or fractions of this value.

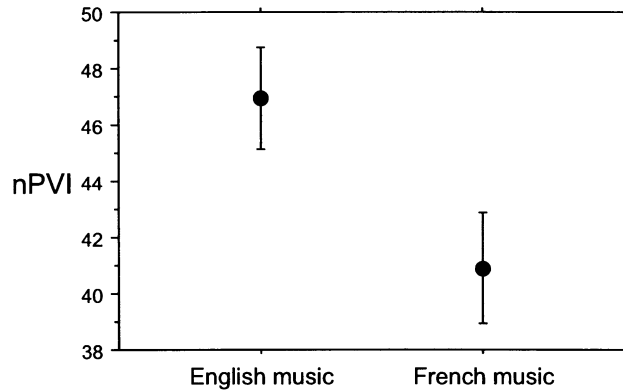


Fig. 3. Musical nPVI values for themes in English and French instrumental classical music (English: mean = 46.91, SE = 1.81, $n = 137$; French mean = 40.90, SE = 1.97, $n = 181$).

possibility we classified all musical themes with respect to meter (binary or ternary), and found that for both cultures approximately 75% of the themes were in binary meter (English: 75.9%, French: 79.0%), and that these proportions were not significantly different between cultures ($P = 0.59$, Fisher's Exact P value). Furthermore, musical themes in binary vs. ternary meters did not differ significantly in their nPVI values (mean of 42.98

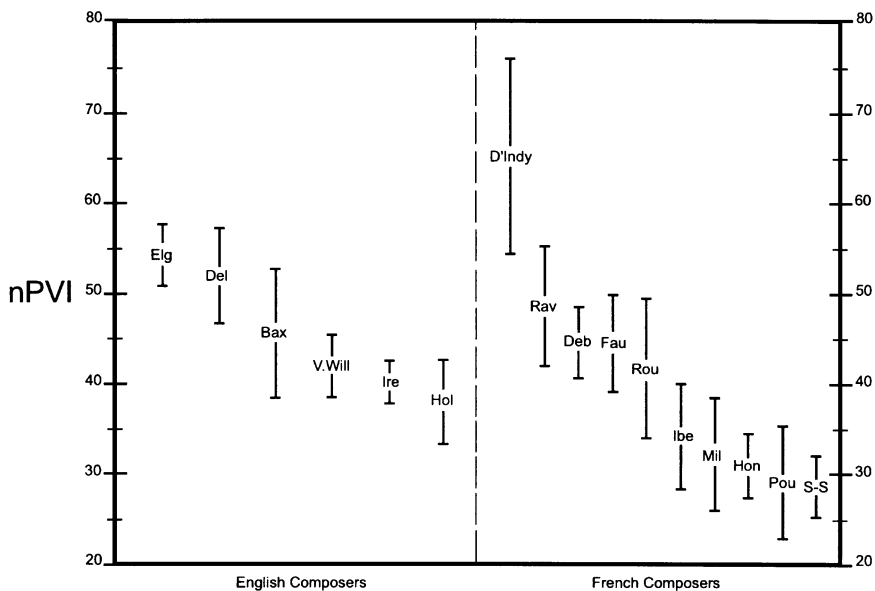


Fig. 4. Musical nPVI values by composer. English composers: Elg = Elgar, Del = Delius, Bax = Bax, V.Will = Vaughan Williams, Ire = Ireland, Hol = Holst. French composers: D'Indy = D'Indy, Rav = Ravel, Deb = Debussy, Fau = Fauré, Rou = Roussel, Ibe = Ibert, Mil = Milhaud, Hon = Honegger, Pou = Poulenc, S-S = Saint-Saëns.

vs. 45.25, respectively, $P = 0.28$, Mann–Whitney U -test). Thus, we can exclude a metrical explanation for the differences in musical nPVI values.

Fig. 4 shows the musical nPVI values for each composer. Among the English composers, Elgar and Delius have the highest nPVI value, while Ireland and Holst have the lowest values. Among the French composers, D'Indy has a very high nPVI value, and is an interesting outlier (see Section 4). The remaining composers range in nPVI from Ravel, Debussy, and Fauré at the high end to Honegger, Poulenc, and Saint-Saëns at the low end. There is a good deal of overlap between the English and French composers in terms of nPVI values, but on average a robust difference emerges, a difference which is in the same direction as the difference in linguistic nPVI values.

4. Discussion

Recent empirical studies of speech rhythm provide the means to address an old question in linguistics and musicology, namely whether the prosody of a composer's native language has an influence on the structure of his or her music. We have applied a quantitative measure of speech rhythm (the nPVI) to the music of 16 composers from two countries which provide prototypical examples of stress- vs. syllable-timed languages: England and France. We find that English and French classical music have significantly different nPVI values, and that this difference is in the same direction as that observed for language (Fig. 3). The observed difference in nPVI values for music is smaller than that for speech, reflecting the within-culture variability of musical nPVI values (Fig. 4). A good deal of intracultural variability is to be expected, however. Music is an artistic endeavor with substantial intracultural stylistic variation and no a priori reason to follow particular rhythmic norms (unlike language). What is remarkable is that in the face of this diversity, average differences still emerge between cultures which parallel the rhythmic differences between the native languages of those cultures.⁵

In conducting this study we focused exclusively on instrumental classical music. Why did we not include classical music written for words, or folk music? The reason is simple: if music is based on words, and words have different rhythmic properties in the languages under study, then it would be no surprise if musical rhythm reflected linguistic rhythm. Indeed, musicologists often attribute one source of musical nationalism around the turn of the last century to the use of folk song melodies in classical music by composers such as Vaughan Williams (Frogley, 1996). We purposely excluded compositions based on songs in order to try to minimize this "obvious" route between language and music. Also, we included many composers whose work is not considered particularly influenced by folk music (e.g. Debussy).

⁵ It is interesting to note that in terms of absolute value, the English musical nPVI is closer to the French linguistic nPVI than it is to the English linguistic nPVI. However, comparison of absolute nPVI values across domains is not meaningful in this study, because the basic units being measured in language and music – vowels vs. tones – tend to have different degrees of sequential variability (e.g. a stressed vowel and a following reduced vowel in speech can differ by an order of magnitude in duration; this degree of sequential contrast in duration is much less common in music). What is meaningful is whether nPVI *differences* within domains are reflected across domains.

If importation of folk music does not explain the differences between English and French musical rhythms, what does? It is known from studies of language acquisition that the perceptual system is sensitive to the rhythmic patterns of language from a very early age (Nazzi et al., 1998; Ramus, 2002a). Composers, like other members of their culture, internalize these patterns as part of learning to speak their native language. One explanation suggests that when composers write music, linguistic rhythms are “in their ears”, and they can consciously or unconsciously draw on these patterns in weaving the sonic fabric of their music. A second explanation proposes that developing composers are influenced by the music of their compatriots. This music, in turn, is influenced by the music these composers hear as children, such as popular and folk songs whose rhythms bear the imprint of their associated linguistic texts. It is important to note that neither explanation suggests that the connection between linguistic and musical rhythm is obligatory: rather, this link is likely to be greater in historical epochs where composers seek a national character for their music.

In this regard, one composer merits special discussion: Vincent D’Indy is a clear outlier in terms of nPVI value compared to other French composers (Fig. 4). Investigation of D’Indy’s background suggests a reason for this pattern. Goldbeck (1965) notes: “[D’Indy] considered French 19th century music ... to be superficial, frivolous and unworthy to compete with the teutonic Bach–Beethoven–Wagner tradition. ... the serious and uncompromising principles of German symphony and of Bayreuth should be applied to French music, inspired by French legends and true to the spirit of French folklore.” If D’Indy’s admiration of German music played a role in his compositional style, and German music reflects the stress-timed German language, then D’Indy’s high nPVI value is understandable. This is speculation, of course, but it does suggest that D’Indy may be an exception that proves the rule rather than a mere outlier.

Future work comparing speech prosody to instrumental music can take at least three directions. First, the nPVI can be used to compare speech and music in a broader variety of languages, including “mora-timed” languages such as Japanese, whose rhythm is distinct from that of English and French (cf. Ramus, 2002; Warner & Arai, 2001). Second, the nPVI can be used to study living composers who speak a single language, in order to see if individual differences in linguistic rhythm (e.g. due to dialectal differences) are reflected in musical rhythm. Third, if empirical differences between the intonation patterns of different languages can be demonstrated, comparison of language and music can be conducted in the melodic domain (cf. Collier, 1991; Hall, 1953/1972; Hirst & Di Cristo, 1998).

Acknowledgements

We thank Esther Grabe and Franck Ramus for providing linguistic nPVI data, and Jennifer Burton, Kathleen Hubbard, John Iversen, Bruno Repp, and two anonymous reviewers for helpful comments. A.D. Patel was supported by an Esther J. Burnham fellowship and by the Neurosciences Research Foundation as part of its program on music and the brain at The Neurosciences Institute.

References

- Abercrombie, D. (1967). *Elements of general phonetics* (p. 97). Chicago, IL: Aldine.
- Abraham, G. (1974). *The tradition of Western music* (pp. 62–83). Berkeley, CA: University of California Press.
- Barlow, H., & Morgenstern, S. (1983). *A dictionary of musical themes*, (Rev. ed.) London: Faber & Faber.
- Besson, M., Faïta, F., Peretz, I., Bonnel, A. -M., & Requin, J. (1998). Singing in the brain: independence of lyrics and tunes. *Psychological Science*, 9 (6), 494–498.
- Carlson, R., Granström, B., & Klatt, D. H. (1979). Some notes on the perception of temporal patterns in speech. *Proceedings of the 9th International Congress of Phonetic Sciences, Copenhagen* (pp. 260–267), 2. Copenhagen: Institute of Phonetics, University of Copenhagen.
- Collier, R. (1991). Multi-language intonation synthesis. *Journal of Phonetics*, 19, 61–73.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51–62.
- Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the 11th International Congress of Phonetic Sciences, Tallinn* (pp. 447–450), 5. Tallinn: Academy of Sciences of the Estonian S.S.R.
- Frogley, A. (1996). *Vaughan Williams studies* New York: Cambridge University Press.
- Goldbeck, F. E. (1965). D'Indy, Vincent. *Encyclopedia Britannica* (p. 452), 7. Chicago, IL: William Benton.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner, *Laboratory phonology* (pp. 515–546). 7. Berlin: Mouton de Gruyter.
- Hall, R. A. (1953). Elgar and the intonation of British English. *Gramophone*, 31, 6. (Reprinted in *Intonation: selected readings*, pp. 282–285, by D. Bolinger, Ed., 1972, Hammondsworth: Penguin)
- Hirst, D. & Di Cristo, A. (1998). *Intonation systems: a survey of twenty languages* Cambridge: Cambridge University Press.
- Jusczyk, R., & Krumhansl, C. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance*, 19 (3), 627–640.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music* (pp. 314–330). Cambridge, MA: MIT Press.
- London, J. (2002). Cognitive constraints on metric systems: some observations and hypotheses. *Music Perception*, 19 (4), 529–550.
- Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language & Speech*, 43 (4), 377–401.
- Nazzi, T., Bertocini, J., & Mehler, J. (1998). Language discrimination in newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24 (3), 756–777.
- Nespor, M. (1990). On the rhythm parameter in phonology. In I. Rocca, *Logical issues in language acquisition* (pp. 157–175). Dordrecht: Foris.
- Palmer, C., & Kelly, M. H. (1992). Linguistic prosody and musical meter in song. *Journal of Memory & Language*, 31, 525–542.
- Patel, A. D. (1998). Syntactic processing in language and music: different cognitive operations, similar neural resources? *Music Perception*, 16 (1), 27–42.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998a). Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience*, 10 (6), 717–733.
- Patel, A. D., Peretz, I., Tramo, M., & Labrecque, R. (1998b). Processing prosodic and musical patterns: a neuropsychological investigation. *Brain & Language*, 61 (1), 123–144.
- Pike, K. N. (1945). *The intonation of American English* (pp. 34–35). Ann Arbor, MI: University of Michigan Press.
- Ramus, F. (2002). Acoustic correlates of linguistic rhythm: perspectives. *Proceedings of Speech Prosody 2002, Aix-en-Provence* (pp. 115–120). Aix-en-Provence: Laboratoire Parole et Langage.
- Ramus, F. (2002a). Language discrimination by newborns: teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition*, 2, 85–115.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292.
- Roach, P. (1982). On the distinction between “stress-timed” and “syllable-timed” languages. In D. Crystal (Ed.),

- Linguistic controversies: essays in linguistic theory and practice in honour of F.R. Palmer* (pp. 73–79). London: Edward Arnold.
- Swain, J. P. (1997). *Musical languages*. New York: W.W. Norton.
- Warner, N., & Arai, T. (2001). The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of America*, 109 (3), 1144–1156.
- Wenk, B. J. (1987). Just in time: on speech rhythms in music. *Linguistics*, 25, 969–981.