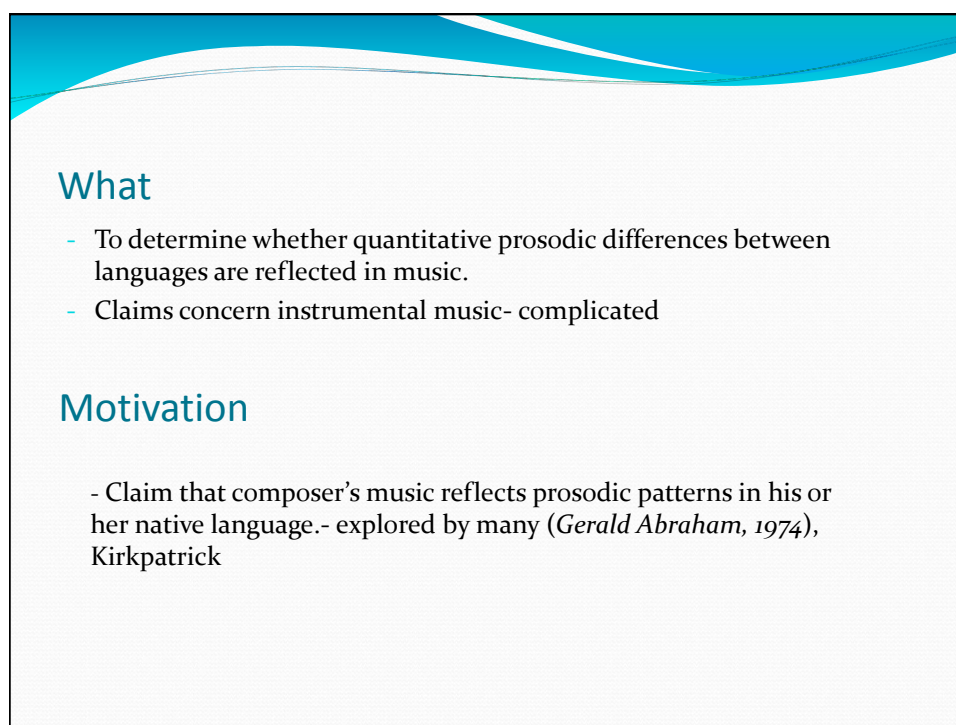


**Comparing the rhythm and melody of speech and music:  
The case of British English and French**  
by Aniruddh D. Patel, John R. Iversen and Jason C. Rosenberg  
The Neurosciences Institute, San Diego, California

Presented by Ravi Nitin Balajee  
USC ISE 575 / EE 675  
January 26, 2010



**What**

- To determine whether quantitative prosodic differences between languages are reflected in music.
- Claims concern instrumental music- complicated

**Motivation**

- Claim that composer's music reflects prosodic patterns in his or her native language.- explored by many (*Gerald Abraham, 1974*), Kirkpatrick

## How???

- The normalized pair-wise variability index (nPVI)- a measure of temporal patterning in speech. (Patel and Daniel, 2003)
  - With regard to rhythm
  - Measures the degree of durational contrast between successive elements in sequence- rhythmic differences between “stress-timed” ,“syllable-timed” and “mora-timed” (something where long syllable consists of two and short syllable consists of one syllable weight)languages .(Low, 1998; Low et al., 2000)(definitions: Pike (1945) and Aber Crombie(1967))
  - nPVI of vowel durations in stress-time languages are higher than syllable timed (Grabe and Low, 2002; Ramus, 2002)- greater degree of vowel reduction in stress-timed languages.(Dauer, 1983, 1987; Nespor, 1990)- similar is the case of duration of notes in instrumental classical themes, proof to compare.

## Tools used

1.

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

- A measure of durational contrast
- Scores range between 0 and 2.
- an absolute number and not a range such as overall variability.
- rhythm based. Can be readily applied to music

-Vowels in English have higher overall durational variability than vowels in French sentences. Hence variability likely to affect nPVI measure

2. Prosogram- quantitative comparison of intonation and musical melody.

- Converts a sentence's fundamental frequency contour (Fo) into sequence of discrete tonal segments- captures perceived pitch pattern of pitch melody.

## Rhythm based studies

- Rhythm- variation of the length and accentuation of a series of sounds
  - Rhythm  $\neq$  isochrony (duration of every syllable is equal). Isochrony might define one kind of rhythm but absence of isochrony is not absence of rhythm!!!
  - Different temporal measures for rhythm
    - $\Delta C$ - Consonantal interval regardless of syllable or word boundary.
    - %V- overall percentage of sentence duration devoted to vowels
- (Ramus *et al.* 1999)
- nPVI
  - Vowel based measures can be easily transferred to music- nPVI chosen because it can be applied to music by measuring durational contrast between successive notes.

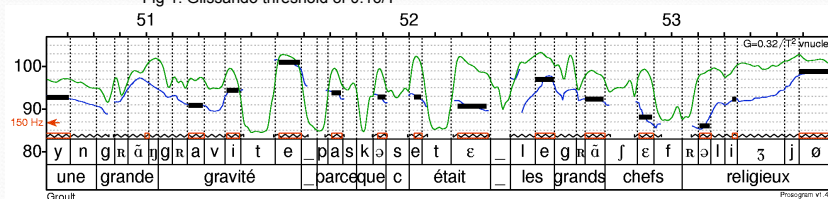
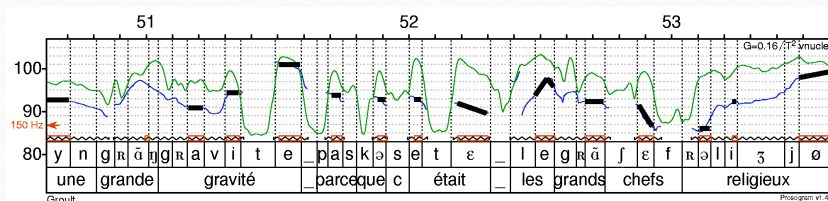
## Ok. nPVI. But is it robust???

- No
- For English and French
  - Ramus(2002)- 67.0 and 49.3
  - Lee and Todd(2004)- 83.9 and 54.3
- English significantly higher than French. Hence considered to be robust.

## Melody based studies

- Pitch variation can be measured by different ways. Variation depends on the method one chooses.
- Use of prosogram.
- Compares the aspect of statistical patterning of spoken intonation.
- Fo stylization principles:
  - Segregation of Fo contour into syllable sized units due to the rapid spectral and amplitude fluctuations in speech signal. (*House, 1990*)
  - Threshold for the detection of pitch movement within a syllable
  - Threshold for the detection of change in slope for pitch movement within a syllable
  - Temporal integration of Fo within a syllable. (*d'Alessandro & Castellengo, 1994; d'Alessandro & Mertens, 1995*)

## Example of Prosogram with stylization variations



Source: <http://bach.arts.kuleuven.be/pmertens/prosogram/#Rationale>

## Prosogram analysis

- Based on segmentation of vowels (or vocalic nuclei) of syllables.
- A glide (or glissando) threshold of  $0.32/T^2$  semitones/second is used (T is the duration of vowel)
  - Choice based on the perceptual research on the threshold for detecting pitch movement in speech
  - Vowels with pitch change below the glide threshold assigned a level tone equal to the median pitch value.

## Methods and criteria:

- Corpora: 20 sentences each in French and English from *Nazzi et al. (1998)*
  - News like utterances, 15 to 20 syllables in length and approximately 3 seconds long.
  - Contains an average of 16 vowels, most of them being singleton.
  - Musical data from turn-of-the 20<sup>th</sup> century English and French composers drawn from musicological sourcebook of instrumental music.
    - Had to meet certain stringent criteria, free from external influences.
    - Resulted in 136 English and 180 French themes

## Phonetic segmentation

- Vowel boundaries marked in English and French sentences using wide-band speech spectrograms generated with SIGNAL defined by Ramus (2002).
- Vowel onset and offset were defined using standard criteria (Peterson and Lehiste, 1960)
- Segmentation of sentences based on display showing the waveform, spectrogram and interactive playback.

## Duration coding of musical themes



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.

- Durations assigned according to time signature
- basic beat assigned a duration of 1. Other notes assigned their relative durations according to standard music notation conventions.
- Flexible- any scheme which preserves the relative duration of notes will result in the same nPVI

Source: *Comparing the rhythm and melody of speech and music*- Aniruddh D. Patel et al.

## Rhythmic and melodic measures and analysis

- The melodic analysis is done using prosogram
- Variability in pitch height and interval is computed via CV
  - For pitch height variation, each level tone assigned a semitone distance from mean pitch of all level tones in sentence then CV is calculated.
  - pitch interval- adjacent level tones are assigned a pitch interval in semi-tones.  $st = \log_2(f_2/f_1)$   $f_1, f_2$ - initial and final tone of pair, CV calculated
- CV(coefficient of variation- measure of overall variability)
  - well suited to compare temporal patterns in speech and music.
  - Sensitive to the relative duration of the events
- Relationship between nPVI and CV computed by 2 ways
  - Linear regression by keeping CV as independent variable and nPVI as dependent variable.
  - Monte-Carlo technique- probability of nPVI difference from existing variability differences- vowel durations in each sentence scrambled and nPVI between 2 languages is computed (1000 iterations)

## Results

### Rhythm :

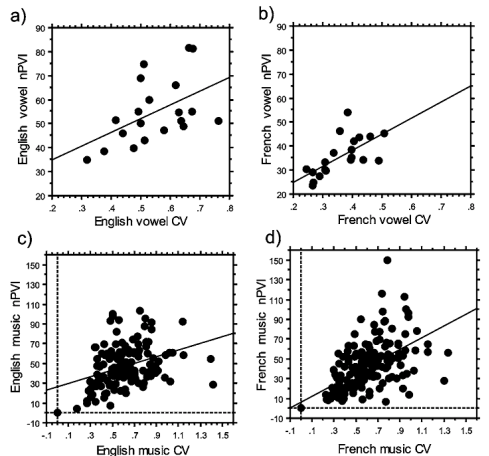
TABLE II. nPVI and CV for speech and music (mean and s.e.). The rightmost column gives the probability that the observed nPVI difference is due to the difference in CV.

	English nPVI	French nPVI	<i>p</i>	English CV	French CV	<i>p</i>	<i>p</i> ( $\Delta$ nPVI  $\Delta$ 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

- Proves nPVI values for English greater than French for both music and speech
- Coefficient of variation (CV) is significant for speech than music

Source: Comparing the rhythm and melody of speech and music- Aniruddh D. Patel et al.

### Regression plot



-higher CV is predictive of higher nPVI for both speech and music, but appears stronger in speech than in music

Source: Comparing the rhythm and melody of speech and music- Aniruddh D.Patel et al.

### Monte-carlo analysis

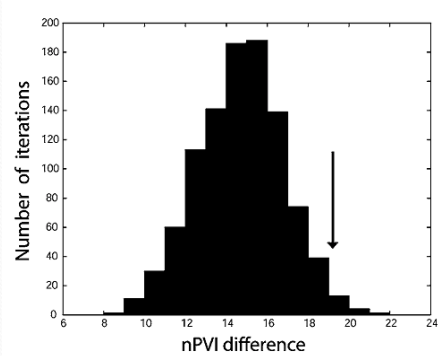


FIG. 4. Result of Monte Carlo analysis for English vs French speech. The actual nPVI difference between the two languages in this study (19.1 points) is shown by an arrow. See text for details.

$p = \Delta nPVI / \Delta CV$  is calculated

- probability of difference greater than 19.1 =  $p = 0.01$  for speech
- For music  $p = 0.001$
- This proves that variability differences is unlikely to account for the nPVI differences in either of the domain.

Source: Comparing the rhythm and melody of speech and music- Aniruddh D.Patel et al.

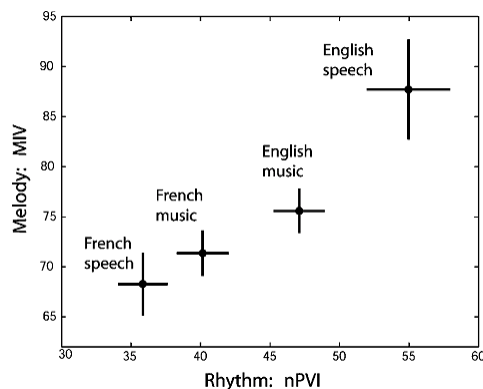
## Melody analysis

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

- English and French speech do not differ in the variability of pitch height within utterances.
- Show a significant difference in pitch interval size variability.
- Shows that as voice moves from one vowel to another, the size of pitch change is more uniform in French speech.
- Similarly for music.

Source: Comparing the rhythm and melody of speech and music- Aniruddh D.Patel et al.

## Rhythm and melody combined



- Melodic interval variability (MIV)= 100\*CV is calculated.
- Position of language in RM space will vary from study to study.
- Distance between the languages=Prosodic distance =

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

- Distance between French and English =27.7 RM units
- Major practical drawback using RM plots- phonetic segmentation of sentences is time consuming.

Source: Comparing the rhythm and melody of speech and music- Aniruddh D.Patel et al.

## Observations and results

- nPVI differences between English and French speech and music are not a by-product of variability differences.
- Music reflects durational contrastiveness in speech and not durational variability.
- Pitch variability reveal that an aspect of speech melody is reflected in music- the variability of interval size between successive pitches in an utterance. But the interval variability between languages is a topic of more research.
- The above tools confirms that music reflects composer's native language.
- Statistical learning of prosodic patters creates knowledge of rhythm and melody patters in speech which inturn can be used to study patterns in music.
  - Direct (reflection of speech in music without influence of folk music) and indirect route (borrowing from folk music bear the stamp of linguistic prosody)of speech patterns into music

## Example

-Hindustani classical and South Indian Classical

<http://www.youtube.com/watch?v=42W5Q5kjmZk>

-Comparison of languages

-South Indian Languages(stress-timed) & Hindi (syllable-timed)

-Hindustani classical- mild influence of Persian & Afgan Music- a little louder and keeps the tone steady, steady beat (Tala, Rhythm)- timbre is pronounced

South Indian music- Influences of cultures in Deep South India-steady beat employed but momentum varies in parts of music to add variations. Timbre is subtle or soft

In the music, the composer chooses Raaga(melody) and Tala (rhythm) and improvises upon them in the performance.

Reference :<http://cnx.org/content/m12502/latest/>

### Future research possibilities

- To examine the relations between Rhythm and Melody to establish patterns and distinguish between music of different nations.
- To find a third dimension in the RM plots that helps clearly distinguish languages.
- To use RM plots to study and distinguish languages from non-native prosody. The prosodic distances can be useful in this regard to find the native language.
  - My perspective- an interesting study would be about rap music.