Acoustics'08
Paris
June 29-July 4, 2008
www.acoustics08-paris.org

# Empirical comparisons of pitch patterns in music, speech, and birdsong

A. T Tierney[a], F. A Russo[b] and A. D Patel[c]

[a]UC San Diego Dept. of Cognitive Science, Neurosciences Institute, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA
[b]Ryerson University Department of Psychology, 350 Victoria Street, Toronto, ON M5B 2K3, Canada
[c]Neurosciences Institute, 10640 John Jay Hopkins Drive, La Jolla, CA 92121, USA
adamtierney@gmail.com

In music, large intervals ("skips") are often followed by reversals, and phrases often have an arch-like shape and final durational lengthening. These regularities could reflect motor constraints on pitch production or the melodic characteristics of speech. To distinguish between these possibilities we compared pitch patterns in instrumental musical themes, sentences, and birdsongs. Patterns due to production-related constraints should be present in all three domains, whereas patterns due to statistical learning from speech should be present in speech but not birdsong. Sequences were taken from classical music of 5 countries, sentences from 4 languages, and songs of 56 songbird families. For sentences and birdsongs each syllable/note was assigned one pitch. For each sequence, we quantified patterns of post-skip reversals, the direction of the initial and final interval, the relative duration of the final vowel/note, and the pitch contour shape. Final lengthening and post-skip reversals predominated in all domains, likely reflecting shared motor constraints; the latter may result from skips' tendency to take melodies toward the edges of the pitch range, forcing subsequent reversals (suggested by Von Hippel & Huron [6]). Arch-like contours were found in music and speech but not birdsong, possibly reflecting an influence of speech patterns on musical structure.

# 1 Introduction

Research over the last decade has identified a number of patterns in the pitch sequences found in a variety of forms of music. Huron[1], for example, found that a large corpus of folk songs exhibited an arch-like shape when phrases of the same note length were averaged together: they tended to contain an initial rise, followed by a plateau, followed by a final fall. Studies have also found that music performers tend to increase the duration of notes just preceding phrase boundaries [2,3], possibly to make the boundaries more clear to listeners. Finally, a much commented-on pattern [4] is that skips—i.e., large jumps in pitch—tend to be followed by reversals in direction.

Despite the interest these patterns have generated, their origins remain obscure. There are at least three possible sources of these regularities; first, they could be due to a conscious effort by the composer to communicate or generate an effect in the listener, as has been claimed for the skip-reversal pattern [4]. Second, they could be due to motor constraints—von-Hippel and Huron [5], for example, present data strongly suggesting that the skip-reversal pattern is due in large part to the fact that pitch distributions in music tend to fall in a roughly Gaussian distribution [6]; large jumps in pitch tend to lead away from the center of this distribution, and thus simple regression to the mean will cause reversals to more commonly follow than continuations.

Finally, another possible source of regularities in musical pitch patterns is statistical learning of pitch patterns present in speech. Like music, speech consists of a sequence of sounds associated with particular fundamental frequency values, and it is perhaps the most prominent feature of one's auditory environment from an early age. Final lengthening, for example, has been found to occur at the boundaries of speech phrases as well [7], especially when there is a syntactic ambiguity that needs to be resolved. It is possible that composers and performers, after hearing final lengthening associated with phrase boundaries in language repeatedly during development, appropriated the same technique when marking phrase boundaries in music. Patel and Daniele [8] presented evidence that another pattern in the rhythm of language—English's tendency to have a larger contrast between neighboring vowel durations, and French's tendency to show a smaller contrast—is reflected in the music of composers from those cultures. Patel *et al.* [9] showed, moreover, that the degree of pitch interval variability of the two languages is also reflected in the music written by composers from each country: both English speech and English music tend to have higher interval variability than French speech and music.

In order to distinguish between these three sources of pattern in music—patterns specific to music, resulting from motor constraints, and learned from speech—we analyzed spoken sentences, birdsongs, and musical themes as sequences of pitches. Any patterns found in all three domains are most likely due to motor constrains. Any patterns found in music and speech, but not birdsong, may be due to statistical learning of speech pitch patterns by composers. Any patterns found in music but not speech may be specific to music, possibly the product of cultural tradition.

# 2 Methods

**Corpora:** The French, English, and Japanese sentences were taken from the database of Nazzi *et al.* [10]. Four female speakers per language read five unique sentences each, for a total of twenty sentences per language. The data set also included three speakers of Yoruba reading seven unique sentences each; these were taken from the database of Marina Nespor and Jacques Mehler. These languages were selected because they possess a wide variety of rhythmic and prosodic features: Yoruba is a tone language, Japanese is a pitch-accent language and is mora-timed, English is a stress-timed intonation language, and French is a syllable-timed intonation language.

Musical themes were selected from Barlow and Morgenstern's *Dictionary of Musical Themes* [11]. Following Patel and Daniele, themes were selected for all English, French, German, Italian, and Russian composers in the dictionary who were born in the 1800s and died in the 1900s. In order to be included, themes were required to contain at least twelve notes and no internal rests, fermatas, or grace notes. Moreover, themes from pieces with titles suggestive of a particular rhythm (e.g. marches, waltzes) or an attempt to produce an exotic style (children's music, music evocative of another composer or country) were also excluded. These criteria yielded 136 English themes from 6 composers, 180 French themes from 10 composers, 112 German themes from 5 composers, 53 Italian themes from 4 composers, and 238 Russian themes from 6 composers.

The birdsong dataset included one song each from 56 of the 84 families of birds in the oscine suborder listed in the *Howard and Moore Complete Checklist of the Birds of the World* [12]. Songs were required to have at least five notes, a tonal quality strong enough for f0 analysis to be performed on them, low background noise, and significant tonal variation (that is, we excluded songs in which only a single note was repeated). All songs consisted of a sequence of notes, both preceded and followed by a long pause, relative to the duration of the notes. Songs were provided by the Cornell Laboratory of Ornithology, the Borror Laboratory of Bioacoustics, the British Museum Library, and compact discs accompanying *Music of the Birds* by Lang Elliot [13], *Nature's Music* by Peter Marler and Hansn Slabbekoorn [14], and *The Singing Life of Birds* by Donald Kroodsma [15].

Duration in musical themes was encoded relative to the time signature, such that the basic beat for each theme was assigned a duration of one. Thus, in the time signature 4/4, a quarter note would be assigned a value of 1, an eighth note would be given a 0.5, etc. Durational data was collected from the speech samples by marking vowel boundaries in each sentence using speech spectrograms generated with Pratt running on a personal computer. Both the waveform and the spectrogram were available during this analysis, plus interactive playback. For birdsong, the onset and offset of each note was marked using wide-band spectrograms generated in SIGNAL running on a modified personal computer (frequency resolution = 125 Hz, time resolution = 8 ms, one FFT every 3 ms, Hanning window).

Pitch sequences in music were encoded as a series of distances from A440. Thus, a note two half-steps above A440 would be encoded as 2, and a note three half-steps below would be encoded as -3. Pitch sequences in speech were encoded using the prosogram version 1.3.6 as instantiated in Pratt. The prosogram is a representation of F0 contour based on human pitch perception. Vowels with a pitch change exceeding the glide threshold $0.32/T^2$ are marked as glides (where T = vowel duration in s). This threshold is based on meta-analysis of a number of studies of the threshold for human perception of pitch change in speech [16]. Vowels with rates of pitch change below this threshold are treated as perceptually equivalent to level tones. Using the threshold $0.32/T^2$ semitones/sec., a large majority of tones are represented as level tones, allowing melodic patterns in speech to be directly compared to music. Pitch sequences in birdsong were encoded using F0 analysis in SIGNAL. The fundamental frequency contour of each note was measured, and the mean pitch was extracted.

In order to further investigate the proposal [5] that skips precede reversals in music due to simple regression to the mean, we calculated the intervals following skips in speech, music, and birdsong. Large jumps in pitch should, more often than not, bring a melody closer to the edge of the available pitch range. Once a melody has landed near the edge of the range, it will most likely reverse direction, for the simple reason that pitches closer to the center of the range are more common than pitches closer to the edge (assuming the pitches fall under a Gaussian distribution). Therefore, if this effect is driving the skip-reversal pattern, skips that cross the median or depart from it should precede reversals, skips that approach the median should be followed by continuations, and skips that land on the median should lead to an equal proportion of reversals and continuations. This was already found to be the case in a corpus of folk songs [5]. Skips (intervals larger than two semitones) were categorized as departing from the median, crossing the median, landing on the median, or approaching the median. The shape of the distribution of pitches in each domain was also assessed by converting pitch values to semitones, then normalizing each note in a given sequence by subtracting the mean pitch of that sequence.

To test the hypothesis that both speech and music show final lengthening, and if so, to question whether or not this is due to motor constraints also shared with birdsong, the durations of the final note of each musical theme and birdsong and the final vowel of each spoken sentence were calculated and compared to all of the other durations within that same domain. A similar comparison was also made between the duration of the initial note/vowel and all remaining notes/vowels.

In order to test the hypothesis that speech, music, and potentially birdsong share the "melodic arch" contour, the initial and final intervals of each phrase was calculated and compared to all of the remaining intervals. In the case of speech, intervals were required to consist of two level tones in order to be included in the analysis.

# 3 Results

Birdsong, speech, and music all showed a tendency for small intervals to predominate over large intervals (figure 1).

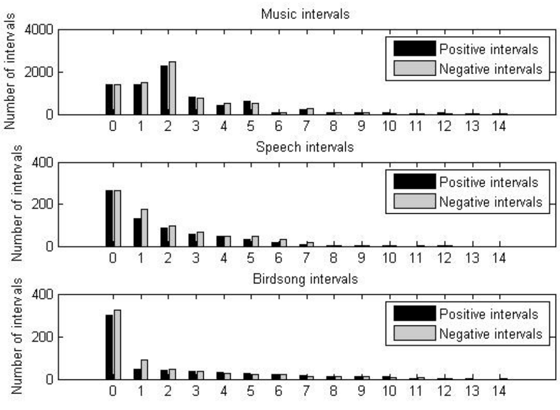


Fig. 1 Interval sizes in speech, music, and birdsong.

Small intervals tend to predominate over large intervals in all three domains, extending to speech and birdsong a finding reported for music by von Hippel and Huron [6]. In addition, a large peak at 2 semitones was found for music but not for speech or birdsong.

Pitches in birdsong, speech, and music fell into a roughly Gaussian distribution, as figure 2 shows.

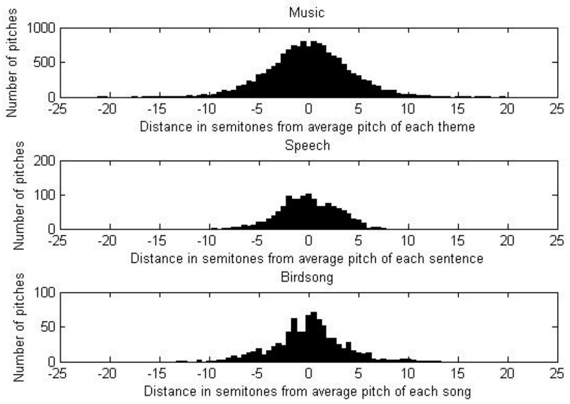


Fig. 2 Histogram of pitches, in distance from average pitch of each song/theme.

This data suggests that pitch sequences in music, speech, and birdsong all show a central tendency, a phenomenon previously observed in music by von Hippel and Huron [6]. As a result, we would expect to find similar skip-reversal patterns in all three domains, as figures 3, 4, and 5 show.

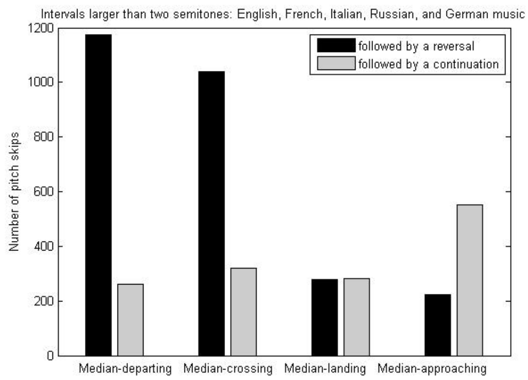


Fig. 3 Skip-reversal patterns in music.

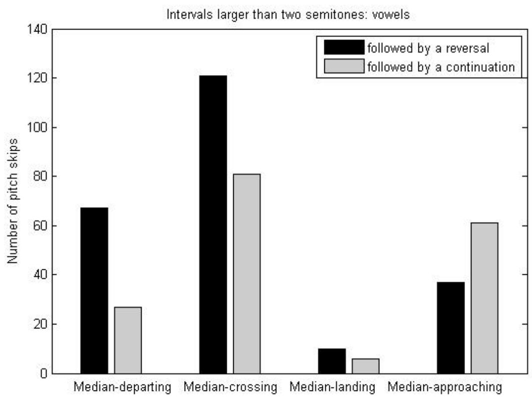


Fig. 4 Skip-reversal patterns in speech.

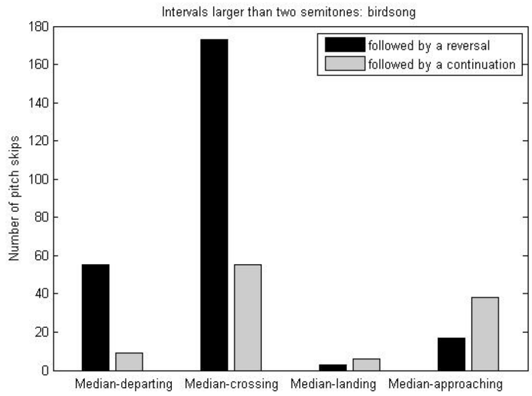


Fig. 5 Skip-reversal patterns in birdsong.

These patterns not only show that skips are followed by reversals in all three domains, but all suggest that in all three cases this is driven at least in part by regression to the mean: median-departing and median-crossing skips tend to be followed by reversals, whereas median-approaching skips tend to be followed by continuations, while median-landing skips do not give rise to a strong pattern.

As figure 6 shows, duration analysis revealed that in speech, music, and birdsong, the last note/vowel tends to be lengthened with respect to the average note. (In each case, the difference between final and average duration was significant.) For music phrases, each beat was arbitrarily given a duration of 50 msec in order to display all domains on the same graph.

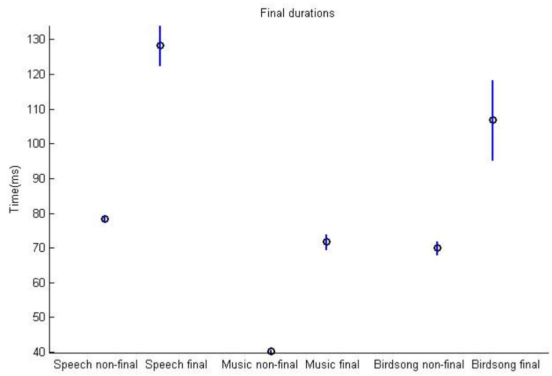


Fig. 6 Final durations.

Figure 7 shows that in speech, the duration of the initial vowel tends to be longer than the average vowel, whereas in music the opposite trend holds. There is no significant difference between the first and average note in birdsong.

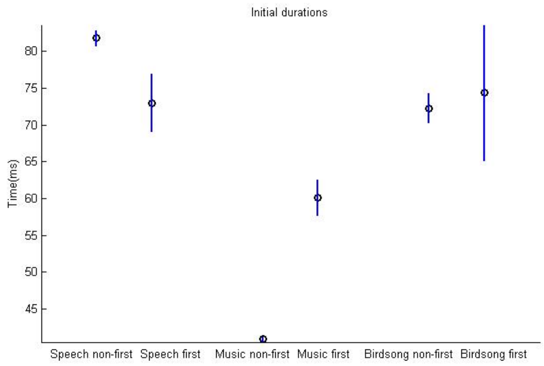


Fig. 7 Initial durations.

As predicted, we found evidence of a "melodic arch" contour in music: final intervals were more negative than average intervals, and initial intervals were more positive than average intervals (figures 8 and 9). Surprisingly, though, this pattern also held true for speech (no significant effect was found for birdsong, and the trend was in the opposite direction).

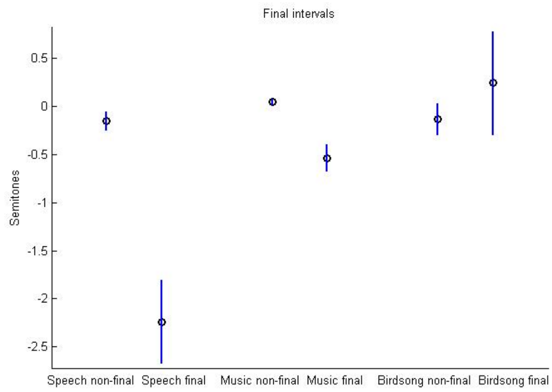


Fig. 8 Final intervals.

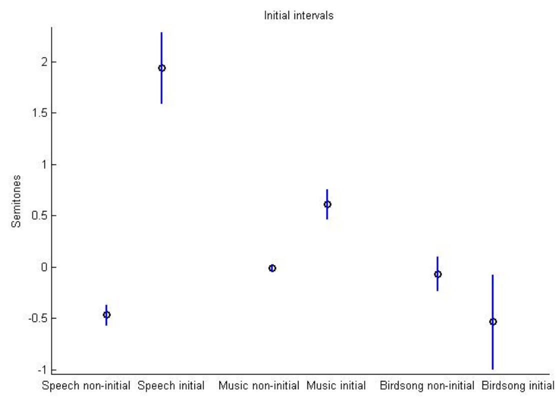


Fig. 9 Initial intervals.

# 4 Conclusion

The present study analyzed spoken sentences, birdsongs, and musical themes for the presence of three patterns: lengthening of notes/vowels at the end of phrases, the presence of an arch-like pitch contour in which pitch rises sharply, plateaus, then falls sharply, and evidence for skips being followed by reversals due to regression to the mean. The latter pattern was found in all three domains, suggesting that it is caused by simple motor constraints: pitches near the center of a speaker/instrumentalist/bird's pitch range are easier to produce than pitches at the edges. This gives rise to distributions displaying central tendency. Thus, the skip-reversal pattern is likely due to regression to the mean in all three cases, and is unlikely to be the consequence of conscious deliberation on the part of either speakers or composers. Final lengthening was also found in all three domains, suggesting that it may be due in part to shared motor constraints, although that does not rule out the possibility that it is taken advantage of by listeners and by speakers as a form of phrase-marking.

A "melodic arch" contour, on the other hand, was found in speech and music, but not in birdsong. This pattern may, therefore, be learned by musicians from the pitch sequences contained in speech, despite the fact that most people are rarely consciously aware of the pitch changes present in speech. If this pattern is indeed shared by both domains, it remains to be determined what function, if any, it serves. It is possible that the effect helps mark the beginnings and ends of phrases, which would facilitate initial syntactic learning in both domains and would help disambiguate ambiguous syntactic structures. It remains to be seen whether listeners are actually able to respond to these cues, and what effect doing so has on their comprehension.

Two patterns were found that appear to be unique to music: a peak in the interval distribution at 2 semitones, and a tendency for initial durations to be longer than subsequent durations. The first effect is most likely due to the tendency of musical melodies to move by small steps rather than leaps [17], a tendency which may itself reflect motor constraints (i.e., smaller intervals are easier to produce than larger ones). Since steps of 2 semitones are more common than steps of 1 semitone in musical scales, this could lead to a predominance of 2-semitone intervals in musical melodies. The cause of the second effect (the tendency for initial durations in melodies to be longer than subsequent ones) is less clear, but it may stem from melodies tending to begin at strong metrical positions.


# Acknowledgments

Supported by Neurosciences Research Foundation as part of its research program on music and the brain at The Neurosciences Institute, where ADP is the Esther J. Burnham Senior Fellow.

# References


[1] D. Huron, "The melodic arch in western folksongs." *Computing in Musicology*10, 3-23 (1996)

[2] B. Repp, "Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists." *JASA* 88, 622-641 (1990)

[3] A. Penel, C. Drake, "Timing variations in music performance: musical communication, perceptual compensation, and/or motor control?" *Perception and Psychophysics* 66, 545-562 (2004)

[4] L. Meyer, *Emotion and Meaning in Music*. University of Chicago Press, Chicago (1961)

[5] P. von Hippel, D. Huron, "Why do skips precede reversals? The effect of tessitura on melodic structure." *Music Perception* 18, 59-85 (2000)

[6] P. von Hippel, "Redefining pitch proximity: tessitura and mobility as constraints on melodic intervals." *Music Perception* 17, 315-327 (2000)

[7] A. Schafer, "Intonational disambiguation in sentence production and comprehension." *Journal of Psycholinguistic Research* 2, 169-182 (2000)

[8] A. Patel, J. Daniele, "An empirical comparison of rhythm in language and music." *Cognition* 87, B35-B45 (2003)

[9] A. Patel, J. Iversen, J. Rosenberg, "Comparing the rhythm and melody of speech and music: the case of British English and French." *JASA* 1995, 3034-3047 (2006)

[10] T. Nazzi, J. Bertoncini, J. Mehler, "Language discrimination in newborns: Toward an understanding of the role of rhythm." *J. Exp. Psychol. Hum. Percept. Perform.* 24, 756-777 (1998)

[11] H. Barlow, S. Morgenstern, *A Dictionary of Musical Themes, revised edition*. Faber and Faber, London (1983)

[12] R. Howard, A. Moore, *A Complete Checklist of Birds of the World.* Macmillan, London (1984)

[13] L. Elliot, *Music of the Birds: a Celebration of Bird Song.* NatureSound Studio, New York (1999)

[14] P. Marler, H. Slabbekorn, *Nature's Music: the Science of Birdsong*. Elsevier, London (2004)

[15] D. Kroodsma, *The Singing Life of Birds.* Houghton Mifflin, New York (2005)

[16] J. 't Hard, R. Collier, A. Cohen, *A Perceptual Study of Intonation.* Cambridge University Press, Cambridge (1990)

[17] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation.* MIT Press, Cambridge (2006)