Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Pitch-interval analysis of 'periodic' and 'aperiodic' Question+Answer pairs

Juan P. Robledo[1], *Sarah Hawkins*[1], *Ian Cross*[1], *Richard Ogden*[2]

[1] Centre for Music and Science, University of Cambridge, UK
[2] Department of Language and Linguistic Science, University of York, UK

jper2@cam.ac.uk, sh110@cam.ac.uk, ic108@cam.ac.uk, richard.ogden@york.ac.uk

Abstract

In English Question+Answer (Q+A) pairs, periodicity typically emerges across turn space, to a degree of precision matching standards of music perception. Interactionally-aligned Q+A pairs display such shared periodicity across the turn, while unaligned pairs do not. Periodicity is measured as temporal location of f0 maxima or minima, 'pikes', in successive accented syllables. This study asks whether periodicity of pikes across a turn is accompanied by systematic use of musical pitch intervals across the turn space. Recordings of 77 Q+A pairs from 8 pairs of native English speakers talking naturally. Ratios of f0 in the last pike of the Question and the first of the Answer fell more reliably into Western musical interval categories when the Q+A pair's turn transition was periodic (the Answer was aligned or preferred, re the Question) than when it was aperiodic (disaligned, dispreferred). Similar results were found for ratios of modal f0. Such pitch ratios are better described by musical interval categories of Western tuning systems than by those of three non-Western systems, and best of all by semitones, suggesting close connections between culturally-specific uses of pitch in conversation and in music. Judgments of arousal/valence suggest weak relations with specific pitch intervals. Theoretical implications are discussed.

Index Terms: f0, conversation, musical pitch intervals, rhythm, alignment, turn-taking, turn space

1. Introduction

As dynamic joint activities in which interactants must coordinate their individual actions to succeed [1], speech and music share rhythm as a central emergent property [2]. By analysing pikes [3] and musical pulse produced by dyads talking while they improvised music, Hawkins, Cross and Ogden [4] concluded that interactants seem to entrain to one another over short periods regardless of domain. Such entrainment in turn seems to provide alignment, understood as gestures or vocalisations produced by the recipient of a conversational turn that are interpretable as supporting the flow of the current activity [5], a finding consonant with views of musical interaction as constitutively aligned [6]. Similarly, Ogden and Hawkins [7] compared instances of rhythmicity vs. non-rhythmicity in Q+A pairs, concluding that rhythmicity is a locally available resource strongly related to social preference, which handles the contingencies of interacting in time by generating interactional alignment and thus facilitating turn-taking. Bearing in mind the musical properties of speech prosody, it is sensible to inquire whether not only rhythm, but also pitch intervals across a turn might be used systematically in conversation. It was hypothesized that, compared with Aperiodic Q+A pairs, Periodic Q+A pairs would exhibit a higher proportion of pitch intervals across the turn space that approximate musical intervals in terms of (Hypothesis 1) the last pike of the question (Q) relative to the first of the Answer (A), and (Hypothesis 2) the mode of the Q relative to those of the A. Furthermore, (Hypothesis 3), the pitch intervals in question would reflect enculturation, and hence the Western tuning system, that being the one familiar to the participants. Hypothesis 2, concerning the mode, came from earlier work on Chilean Spanish [8].

2. Method

2.1. Participants and dataset

Data comes from eight same-sex pairs of friends (four pairs each sex) aged 19-31 years (mean 24) recorded at Cambridge University [4]. All were university educated, native speakers of British English (England, Scotland, Northern Ireland).

2.2. Rhythm analysis

77 of the Q+A pairs produced were analysed for pitch. Pike timing was measured by hand using Praat's f0 tracker with default settings (v. 5.3.19). Following [7], the 77 Qs were classed as rhythmic when the time intervals between the last 3 successive pikes differed by no more than ±15%. Arrhythmic *Qs* display no such periodicity of pikes. After a rhythmic *Q*, entry into the turn space was defined as periodic when the first pike of the A came in on the beat established by the *Q's* pikes, thus generating a periodic Q+A pair. Resultant classifications of Q+A pairs were: 34 (44%) Periodic, 25 (32%) aperiodic, and 18 (23%) arrhythmic Qs.

2.3. Pitch analysis

2.3.1. Local measurement: last:first pair pike ratio

The f0 of each Q's last pike and the first pike of its A were measured and their ratio (last:first pike ratio) was calculated in cents (1 cent is 100^{th} of a semitone) using the formula:

$$Cents = 1200*log2(f1/f2) \quad (1)$$

where *f1* and *f2* = f0 of the last Q and first A pikes, *f1* being the higher and *f2* = the lower f0.

The jnd can be about 5 cents, but pitch differences of 20-25 cents are usually identified as 'in tune' [9, 10].

The 3 cases for which ratios were larger than an octave were converted to their equivalents within the one octave, as is customary in the conceptualization of pitch class hierarchies in tonal music. For 9 tuning systems and scales, we classed pitch intervals as *musical* or *indeterminate.* For the 12 semitones of the equal temperament (ET) chromatic-scale octave, intervals classed as musical fell at the theoretical ratio for a particular

 doi: 10.21437/SpeechProsody.2016-220

musical interval ±25 cents in one analysis, and ±15 cents in a second. Indeterminate intervals fell outside these pitch ranges. Thus the ±25 cent analysis sorted all intervals into 50-cent bins, yielding a 50% chance of any given ratio being classed as musical. To illustrate using Table 1, for a last:first pike ratio of 9/8—a major second—intervals 175-225 cents above the unison (1:1) ratio set by fA were classed as a major second (musical); while intervals within 150-174 or 226-250 cents were indeterminate, associated with minor and major seconds respectively. In the ±15 cent analysis, the chance of being classed as musical was 30%, and of indeterminate 70%.

The ET chromatic scale comprises only one set of values for pitch interval categories. Other Western intervallic systems explored were Just intonation (JI) and the Western seven-step major and natural minor diatonic scales, in both ET and JI, all with 50-cent bins. See [11, 12] for relevant explanations. Table 1 shows, for JI and the intervals of the Western ET major scale, that cent step sizes differ for each tuning system and scale, with commensurate differences in the chance of any observed interval being classed as musical or indeterminate.

Step #	Chromatic intervals (common names)	Cents		
		ET	JI	ET Maj
0	Unison (Un)	0	0	0
1	Minor second (m2)	100	112	-
2	Major second (M2)	200	204	200
3	Minor third (m3)	300	316	-
4	Major third (M3)	400	386	400
5	Perfect fourth (P4)	500	498	500
6	Tritone (TT)	600	590	-
7	Perfect fifth (P5)	700	702	700
8	Minor sixth (m6)	800	814	-
9	Major sixth (M6)	900	884	900
10	Minor seventh (m7)	1000	1018	-
11	Major seventh (M7)	1100	1088	1100
12	Octave (Oct)	1200	1200	1200

Table 1. *Western ET and JI tuning systems expressed through their absolute step number, name, and distance in cents from a constant pitch (0 cents). Adjacent step numbers represent semitone (100 cent) differences in ET, but smaller or larger numbers of cents in other tuning systems/scales, as shown for the 12 steps of JI chromatic, and the 7 steps of ET major scale.*

To test for the cultural biases predicted by Hypothesis 3, the observed pitch intervals were compared with the tuning systems of three non-Western scales: Pélog, Slendro (both Javanese) and Chopi (Portuguese East Africa). The Western major and minor diatonic scales and the three non-Western scales comprise a smaller number of steps between octaves than the Western chromatic scale, and hence a majority of step sizes larger than 100 cents. As such, 50-cent bins would result in the majority of ratios being classed as indeterminate, e.g. 79% in the case of Slendro, 75% for Chopi and 70% for Pélog. To give a 50% chance of a ratio being classed as musical, alternative thresholds were introduced for each non-Western scale, proportional to each one's average step size. This was in addition to 50-cent bins. These larger bin sizes were ±60 cents (120-cent bins) for Slendro and ±43 cents (86-cent bins) for Pélog and Chopi. This procedure works against the hypothesis of enculturation, since the criterion for labelling a prototype is more lax for the non-Western scales [13].

2.3.2. Non-local measurements: mode:mode ratios

Following [8], the f0 modes (Hz) were calculated from the complete f0 contour of each Q, and of its corresponding A (using a Praat script), and a 'mode:mode' ratio calculated for each Q+A pair. Analyses used the same cents and octave conversions, interval categorization and bin sizes, as before.

3. Results

3.1. Correlation of f0 range with last:first pike ratio

To assess whether the Q's f0 range might influence last:first pair pike ratios (e.g. in case an answer reflected a wide f0 range in its question), the correlation across all Q+A pairs was calculated between the Q's f0 range (from the whole f0 contour, not just the pikes) and each Q+A pair's last:first pair pike ratio size. There was essentially no relationship (Spearman's Rho (skewed) = -0.03, n = 77, S = 13668, p = 0.84). It was concluded there was no such effect.

3.2. Local measurements: last:first pair ratios

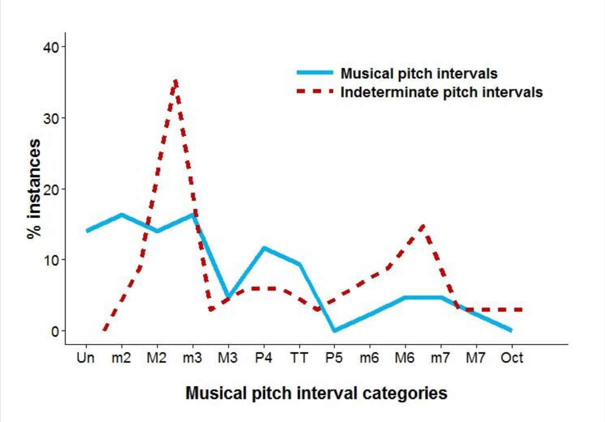

Figure 1: *Percentage distribution of last:first pair pike ratios within or outside ±25 cents of the ideal ratio of each interval in the equal-temperament chromatic scale. Total n = 77. Blue solid curve: ratios that fall within ±25 cents of an ideal interval (musical). Red dashes: ratios that fall outside the ±25-cent limit (indeterminate intervals).Though the data are binned into discrete categories, they are shown as curves for visual clarity. See Table 1 for category names (e.g. m2, M2).*

Figure 1 shows the percentage distribution of last:first pair pike ratios, each assigned to one of the 12 musical intervals of the ET chromatic scale: musical intervals (±25 cents) at the interval on the axis, and indeterminate intervals between marked interval sizes. The distributions are similar in that instances of smaller intervals outnumber instances of larger ones, unsurprisingly, but they are quite different in other respects. Most notable is a predominance of indeterminate intervals between M2 and m3 (35%) and M6 and m7 (15%), and complete absence of P5 and rarity of M3 amongst musical pitch intervals. Neither rhythmicity of the Q nor periodicity vs aperiodicity of the A seemed to influence these patterns.

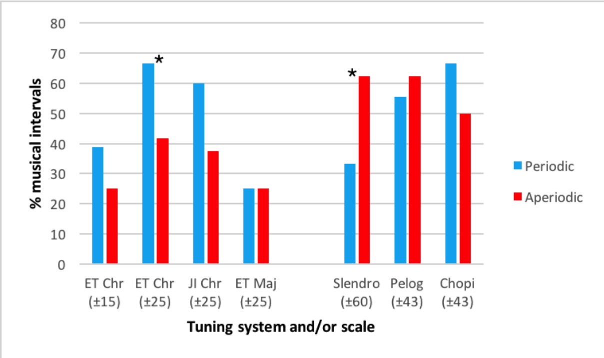

Figure 2: *Percentage of last:first pair pike musical intervals relative to the total (musical+indeterminate) for Periodic (blue) and Aperiodic (red) Q+A pairs, for some tuning systems and scales. ET Chr: equal-temperament chromatic. JI Chr: just-intonation chromatic. ET Maj: equal-temperament major diatonic. Bin sizes (cents, in parentheses) adjusted so chance of an interval being classed as musical or indeterminate is 50%, except for ET Chr ±15 cents (far left), where chance = 30% for musical and 70% for indeterminate intervals, and for ET Maj, where ±25 cents is 50% for 2 intervals but 25% for 5 others. * p = 0.056 (ET chr); p < 0.03 (Slendro).*

To assess Hypothesis 1, the percentage of musical (as opposed to indeterminate) intervals was calculated for the 34 periodic and 25 aperiodic Q+A pairs separately, for each tuning system, scale, and bin size used in the study. Figure 2 shows the data for the three Western systems shown in Table 1 and all three non-Western analyses; bin sizes (two for ET chromatic) are on the x-axis. In the Western chromatic and minor scale analyses, there was a higher proportion of musical intervals across the turn space in periodic than in aperiodic Q+A pairs, although the difference only approached significance for chromatic scales with 50-cent bins: ET chromatic 67% vs 42%, $\chi^2(1) = 3.67$, $p = 0.056$; JI chromatic 60% vs 38%, $\chi^2(1) = 3.12$, $p = 0.073$). Two of the three non-Western systems showed the opposite tendency (Chopi did not), but the difference was only significant for Slendro with 120-cent bins ($\chi^2(1) = 4.95$, $p = 0.03$). In all other cases, Western and non-Western, differences between periodic and aperiodic contexts fell far short of significance regardless of bin size ($p > 0.1$).

Hypothesis 1, that the last:first pike analysis would show a higher proportion of musical intervals across the turn space, is thus weakly supported for Western chromatic scales, and not at all for the other pitch-interval systems explored.

3.3. Non-local measurements: mode:mode ratios

Fig. 3 shows the same analysis as Fig. 2, but for mode:mode intervals rather than last:first pike. Both ET and JI chromatic scales show the same pattern as the last:first pike analysis: more musical than indeterminate intervals in periodic than aperiodic Q+A pairs, with the two categories spanning the 50% chance level; and this time each achieves statistical significance: ET Chr ±25 cents: 65% vs 33%, $\chi^2(1) = 5.35$, $p = 0.02$; JI Chr 62% vs 33%, $\chi^2(1) = 4.46$, $p = 0.035$). ET Chr ±15 cents was marginally significant ($\chi^2(1) = 3.47$, $p = 0.06$). All other tuning systems and bin sizes fall far short of significance ($p > 0.1$), although unlike the first:last pike analyses, in the mode:mode analyses the pattern of more musical intervals in periodic than aperiodic Q+A pairs is consistent across all Western systems.

In sum, the mode:mode results support Hypothesis 2 more strongly than the last:first pike analyses support Hypothesis 1,

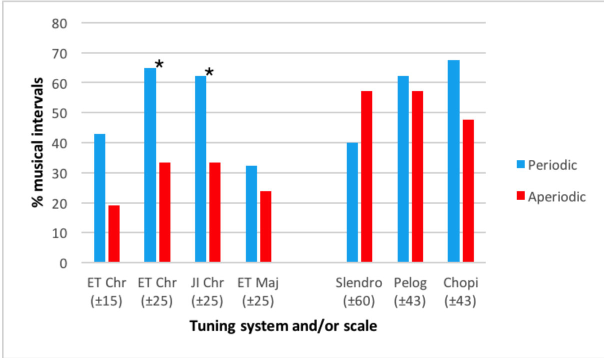

Figure 3: *Same as for Figure 2 except the data are for the mode:mode analysis as described in the text. * = p < 0.04.*

and the overall mode:mode picture is more consistent. Both last:first pike and mode:mode analyses support Hypothesis 3, that, to the extent that musical intervals appear across the Q+A turn space, those intervals are likely to be those of the prevailing musical culture. Specifically, they are tuned to the semitone-based intervals of the Western chromatic scale, though they include avoidance of some particular intervals.

3.4. Associations with arousal and valence: initial analyses

To see if specific pitch intervals are associated with different types or degrees of emotion, the first 2 authors independently judged arousal and valence for each Q and each A (including arrhythmic Qs and their As) on a 9-point scale. They used auditory and visual cues, with minimal reference to lexis and grammar. They 'trained' on 10 Q+A pairs, reaching close agreement, then scored the rest independently. On 34 items, intra-judge reliability for the 2nd author was high ($r = 0.9$ for each of the 4 measures); inter-judge reliability was acceptable (arousal: $r = 0.6$ for Q, 0.65 for A; valence: $r = 0.7$ for both Q and A, $p << 0.001$ in each case). Encouraging though these numbers are, observations are tentative for 3 reasons: criteria are hard to use and may not be optimal; the conversational styles meant the full range of emotions was rarely expressed, with strong negativity especially rare; and the data are sparse, especially for semitone analyses. Thus we make only tentative general observations here, without supporting statistics.

With respect to the musical and indeterminate intervals shown in Fig. 1, no explanation has yet been found, not even for the frequency of indeterminate intervals between M2-m3, and M6-m7. However, valence was higher in Answers that had musical rather than indeterminate pitch intervals. Perhaps greater positivity is associated with more musical talk. Considering now rhythmicity and affect (Figs. 2 & 3), periodic Answers tend to follow questions that were asked with less arousal; aperiodic Answers tend to have lowest valence; while answers to arrhythmic Questions tend to have highest valence.

4. Discussion

Our data show that, analysed on semitone scales, periodic As employ more musical intervals than aperiodic As do across a turn, both on pikes that span the turn, and more strongly in the case of the modal f0 of the Q relative to that of the A. Further, periodic As exhibit more musical than non-musical intervals overall, whereas aperiodic As predominantly use indeterminate pitch intervals across turns.

The patterns observed suggest that a relatively high degree of musicality accompanies preferred social actions: not just greater rhythmicity across turns, but also more musical intervals used when the Q+A pair is periodic (i.e. rhythmic), as well as, tentatively, more positivity (higher valence). This suggests a high degree of acoustical, musical and emotional coherence when responses are preferred or aligned, and various ways of reducing such coherence when they are less preferred, or less well aligned. For example, preferred As may be more likely to enter the turn space using the rhythmic beat set up by a well-formed Q, using a pitch that falls close to a musical interval, and more positive affect (valence). In contrast, even when the Q sets up a rhythmic beat, an A that is dispreferred may signal its non-alignment by starting off-beat, using a pitch that does not approximate a musical interval, and conveying a less positive affective stance through tone of voice and gesture. Interestingly, periodic Q+A pairs seem to be associated with low arousal, whereas arrhythmicity in a Q may be greeted with a highly positive A (high valence). This suggests that arrhythmicity may be another way of producing affiliative responses, perhaps during heightened affect.

The better match of our data to the ET chromatic scale than to other tuning systems, Western and non-Western, reflects the fact that the ET chromatic scale is the only one of those tested to use a consistent, and relatively small, step size, as explained in the Introduction. However, the fact that the difference between periodic and aperiodic Q+A pairs was significant for 50-cent but not 30-cent bins show that small step size is not the only influential factor. Rather, our talkers were attuned to the 12-step semitone scale, and to deviations from its ideal pitches roughly in line with what Western listeners judge as 'in' vs. 'out' of tune. The other scales and tuning systems did not differentiate so well simply because their step sizes were either coarser (the majority) or finer (ET chromatic with 30-cent bins) than the degree of precision talkers used.

The absence of P5 (perfect 5th) intervals amongst periodic Q+A pairs is interesting since, after the tonic, P5 is arguably the most important interval in the diatonic scale, which forms the basis of Western harmony. M3 (major 3rd) is almost as important in this respect, and is also relatively rare in our data. It is as if these basic harmonic relationships are deliberately avoided between interlocutors as they talk. The scarcity of M3 but not m3 is also noteworthy since use of a minor 3rd *within* an utterance is widespread as a 'calling contour', at least across European languages [14]. Further, M3 and P5 predominate in infant-carer vocal interaction [11].

It is tempting to suggest that P5 is avoided between adult interlocutors because it represents a 'pivot' region, above which pitch intervals between interlocutors are reacted to as marked (unusual), and below which they are more typical. Given that most talkers lack the precise pitch perception of trained musicians, and presumably need only approximately match pitches with fellow talkers, avoidance of entire interval regions is an appealing notion. It conforms to [15]'s proposal that 'innominate' regions of the colour spectrum can play powerful roles in memory and hence cognitive organization. However, if P5 is such a pivot region, one would expect raised affect with larger intervals, for which we found no evidence in our data, although that may be due to the somewhat homogeneous nature of our dataset, as explained below.

Our findings confirm and extend the literature demonstrating fine-grained interpersonal pitch entrainment, described as *tonal* [11] or *pitch* [16] *synchrony.* Contextualised within the functionally-defined framework of Conversation Analysis, our findings show that its use is partially governed by the role of the conversational turn, and tied relatively straightforwardly to rhythm. Relationships with affect seem more complex: our measures to date conform with general expectations but offer little insight into details of our observed patterns. Much more work is needed in this area, including, presumably, analyses of lexis and grammar, as well as systematic analyses of gesture. Our current data set may be unsuited to deeper analysis, since the affect displayed is relatively homogeneous: though some talkers were highly emotive, all conversations were amicable, there were no arguments, and very little negative emotion expressed. This is at least partly due to our design, especially our deliberate encouragement of cooperative tasks. Future work should encourage a wider range of tasks. The work of [8], showing for Chilean Spanish that f0 mode:mode matching is closer when talkers trust one another, indicates that such work would be worthwhile.

The stronger relationship between periodicity and pitch intervals for the mode:mode than the pike analyses might suggest that this is because pitch is managed more globally than locally. However, this seems unlikely. First, in a small data set like ours, modes may be more statistically stable than last:first pikes, given the wealth of other influencing variables. Second, vocal accommodation to average pitch is also common, and will presumably affect more local processes. Third, closer pitch matching between talkers has been shown for aligned insertions (e.g. *uh-huh*) compared with non-aligned ones, and these short utterances are said to be locally managed [17]. These authors also discuss the variety and complexity of potential influences on pitch alignment. However, other work does claim non-local dependencies in musical syntax and key relationships [18, 19]. Presumably local and global influences both exist, and their complexities are such that it will be some time before we can describe them well, let alone explain them.

Taken together, these rhythmic and musical pitch interval findings suggest a close connection between prosodic and musical behaviour. Though this suggestion is not new (speech prosody has long been seen as musical, and cf. [4, 20, 21]), this paper offers insights into how these close connections work. Both pitch-interval and rhythm measures operate across rather than just within turns, indicating the value to the science of human communication of a grammar and phonology that can span utterances and turns between talkers. They work together: pragmatically-aligned interactions involve more strongly-entrained rhythm and more precise enculturated pitch intervals; disaligned/dispreferred interactions do not. This conclusion is supported by neuroscientific evidence that reduced temporal variation facilitates joint action and enhances attention [22-24], that brain activity synchronizes during musical and social interaction alike, and is almost certainly domain-general [25-2825-28], and that these fleeting, on-the-fly mutual accommodations seem essential for real-time interpersonal communication, be it spoken or musical [29-31]. Lastly, mutual accommodation to pitch intervals is self-evidently learned by cultural transmission, but temporal entrainment seems likely to be a basic biological phenomenon essential to all types of cooperative joint action.

5. Acknowledgements

Supported in part by the Cambridge Commonwealth, European & International Trust, British Academy Grant SG120400 and the Cambridge University Newton Trust.

6. References

[1] Borrie, S. A. & Liss, J. M. (2014). Rhythm as a Coordinating Device: Entrainment With Disordered Speech. *Journal of Speech, Language and Hearing Research (Online)* 57(3), 815-824.

[2] Hausen, M., Torppa, R., Salmela, V. R., Vainio, M. & Särkämö, T. (2013). Music and speech prosody: A common rhythm. *Frontiers in Psychology* 4(566). [10.3389/fpsyg.2013.00566]

[3] Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture* 7(2), 179-214. [10.1075/gest.7.2.04loe]

[4] Hawkins, S., Cross, I. & Ogden, R. (2013). Communicative interaction in spontaneous music and speech. In *Language, Music and Interaction*. Orwin Martin, Howes Christine, Kempson Ruth (Eds.) pp. 285-329. London: College Publications.

[5] Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language & Social Interaction* 41(1), 31-57. [10.1080/08351810701691123]

[6] Cross, I. (2013). "Does not compute"? Music as real-time communicative interaction. *AI and Society* 28(4), 415-430.

[7] Ogden, R. & Hawkins, S. (2015). Entrainment as a basis for co-ordinated actions in speech. In: 2015 The Scottish Consortium for ICPhS, (Ed.). 18th International Congress of Phonetic Sciences. Glasgow, UK: The University of Glasgow; ISBN 978-0-85261-941-4. Paper number 0599.

[8] Robledo, J. P., Hurtado, E., Prado, F., Román, D. & Cornejo, C. (accepted). Music intervals in speech: An interpersonal approach. *Psychology of Music*.

[9] Vurma, A. & Ross, J. (2007). Timbre-induced pitch deviations of musical sounds. *Journal of Interdisciplinary Music Studies* 1(1), 33-50.

[10] Curtis, M. E. & Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion* 10(3), 335-348. [10.1037/a0017928]

[11] Van Puyvelde, M., Vanfleteren, P., Loots, G., Deschuyffeleer, S., Vinck, B., Jacquet, W. & Verhelst, W. (2010). Tonal synchrony in mother–infant interaction based on harmonic and pentatonic series. *Infant Behavior and Development* 33(4), 387-400. [http://dx.doi.org/10.1016/j.infbeh.2010.04.003]

[12] Van Puyvelde, M., Loots, G., Gillisjans, L., Pattyn, N. & Quintana, C. (2015). A cross-cultural comparison of tonal synchrony and pitch imitation in the vocal dialogs of Belgian Flemish-speaking and Mexican Spanish-speaking mother–infant dyads. *Infant Behavior and Development* 40, 41-53. [http://dx.doi.org/10.1016/j.infbeh.2015.03.001]

[13] Perlman, M. & Krumhansl, C. L. (1996). An experimental study of internal interval standards in Javanese and Western musicians. *Music Perception: An Interdisciplinary Journal* 14(2), 95-116. [10.2307/40285714]

[14] Ladd, D. R. (2008). *Intonational Phonology*. Cambridge: Cambridge University Press.

[15] Berlin, B. & Kay, P. (1991). *Basic Color Terms: Their Universlity and Evolution*. Berkeley: University of California Press.

[16] Okada, B. M., Lachs, L. & Boone, B. (2012). Interpreting tone of voice: Musical pitch relationships convey agreement in dyadic conversation. *The Journal of the Acoustical Society of America* 132(3), EL208-EL214. [doi:http://dx.doi.org/10.1121/1.4742316]

[17] Gorisch, J., Wells, B. & Brown, G. J. (2012). Pitch contour matching and interactional alignment across turns: An acoustic investigation. *Language and Speech* 55(1), 57-76. [10.1177/0023830911428874]

[18] Koelsch, S., Rohrmeier, M., Torrecuso, R. & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences* 110(38), 15443-15448. [10.1073/pnas.1300272110]

[19] Woolhouse, M., Cross, I. & Horton, T. (2015). Perception of nonadjacent tonic-key relationships. *Psychology of Music*, 1-14. [10.1177/0305735615593409]

[20] Levinson, S. C. & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 369(1651). [10.1098/rstb.2013.0302]

[21] Cross, I. (2012). Music and biocultural evolution. In *The Cultural Study of Music: A Critical Introduction*. Clayton Martin, Herbert Trevor, Middleton Richard (Eds.), 2nd ed, pp. 17-27. London: Routledge.

[22] Arnal, L. H. & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *TRENDS in Cognitive Sciences* 16(7), 390-398. [10.1016/j.tics.2012.05.003]

[23] Lakatos, P., Shah, A. S., Knuth, K., H, Ulbert, I., Karmos, G. & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology* 94, 1904-1911. [10.1152/jn.00263.2005]

[24] Vesper, C., van der Wel, R. P. R. D., Knoblich, G. & Sebanz, N. (2011). Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research* 211, 517-530.

[25] Müller, V., Sänger, J. & Lindenberger, U. (2013). Intra- and inter-brain synchronization during musical improvisation on the guitar. *PLoS ONE* 8(9), e73852. [10.1371/journal.pone.0073852]

[26] Dumas, G., Nadel, J., Soussignan, R. & Martinerie, J. (2010). Inter-brain synchronization during social interaction. *PLoS ONE* 5(8), e12166. [10.1371/journal.pone.0012166]

[27] Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S. & Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *TRENDS in Cognitive Sciences* 16(2), 114-121. [10.1016/j.tics.2011.12.007]

[28] Cason, N., Astésano, C. & Schön, D. (2015). Bridging music and speech rhythm: Rhythmic priming and audio–motor training affect speech perception. *Acta Psychologica* 155, 43-50. [http://dx.doi.org/10.1016/j.actpsy.2014.12.002]

[29] Wilson, M. & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review* 12(6), 957-968.

[30] Pickering, M. J. & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36, 329-392. [10.1017/S0140525X12001495]

[31] Sänger, J., Müller, V. & Lindenberger, U. (2012). Intra- and interbrain synchronization and network properties when playing guitar in duets. *Frontiers in Human Neuroscience* 6. [10.3389/fnhum.2012.00312]