Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

Comparing Models of Melodic Contour in Music and Speech

Alex Billig,[*1] Daniel Müllensiefen[*2]

[*]*Department of Psychology, Goldsmiths, University of London, United Kingdom*

[1]ajbillig@gmail.com, [2]d.mullensiefen@gold.ac.uk

ABSTRACT

Background

Contour, the pattern of rises and falls in pitch of a sequence of sounds, is an important perceptual and mnemonic feature of both music (e.g. Dowling & Bartlett, 1981) and speech (e.g. Frazier, Carlton, & Clifton, 2006). Various models have been put forward to formally define contour in music, varying in the degree to which they compress the melodic information. However, it is unclear whether any of these uses a level of abstraction comparable to that involved in the mental processes underlying melody perception and memory.

In contrast to the discrete pitch steps typical of most musical melodies, the fundamental frequency of a speech utterance tends to vary continuously. Much intonation research has focused on the extent to which this variation results in perceived changes in pitch (Hermes, 2006). However, little is known about memory representations of speech pitch patterns. Kochanski (2009) suggested that although discrete contour categories may play a role in memory for intonation, at least some of the finer pitch detail is also retained over short periods. The present study uses musical models for the first time to probe the cognitive processes underlying contour processing in speech.

Finally, a growing body of literature suggests that musical training is associated with neural changes in pitch processing of both musical notes (e.g. Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004) and spoken syllables (e.g. Schön, Magne, & Besson, 2004). However, whether such experience results in qualitative differences in the mental representation of longer pitch patterns (whole melodies or sentences) is less clear, and a question also addressed by this study.

Aims

The aims of the study were: (1) To evaluate the cognitive adequacy of four contour models, (2) To compare perception and memory for contour across the music and speech domains, (3) To investigate the degree to which mental representations of pitch trajectories vary with musical experience.

Method

Thirty-two participants with a range of musical experience listened to a series of short monophonic melodies (taken from a large database of popular music) and spoken English sentences (low-pass filtered at 400 Hz to remove semantic information). After each auditory presentation they were asked to identify which of four visual images best represented the melody or sentence heard. All images in a single trial were produced using the same contour model (from a selection of four, described below) but only one of the four images was derived from the auditory stimulus.

Four different contour models were compared: (1) Huron's simple contour (Huron, 1996), which displays only the relationship (same/higher/lower) between the pitch of the first note, the pitch of the last note, and the average pitch of all the notes in between, (2) Curve contour (Müllensiefen & Wiggins, 2011), in which a polynomial curve is fitted to the note data (pitch and onset time) using least squares regression, (3) Reduced Line contour, which represents all note onsets precisely but only uses two levels of pitch interval (smaller steps and larger skips), (4) Step model contour (e.g. Eerola & Toiviainen, 2006), in which each note is plotted based on its exact pitch and onset time. In order to create the images for the speech stimuli, these were first converted into a series of tonal segments using the Prosogram (Mertens, 2004) script in the speech processing software Praat.

The experiment comprised eight blocks: one for each combination of contour model and stimulus type (melodies and sentences). Each block consisted of three practice trials followed by sixteen experimental trials. The order of contour models and stimulus types was counterbalanced across participants, and the order of stimuli in a block randomised for each participant.

The premise of the paradigm is that contour models facilitating the highest proportion of correct matches can be considered to summarise the pitch information in a cognitively optimal way. The cross-modal approach draws on evidence that listeners, regardless of musical experience, can reliably interpret pitch trajectories as visual shapes (e.g. Balch & Muscatelli, 1986).

Musical training was assessed using the relevant sub-scale of the Goldsmiths Musical Sophistication Index (Müllensiefen, Gingras, Stewart, & Musil, 2011).

Results

Binomial tests revealed that participants selected the target image significantly more frequently than at chance level for all models for both melodies and speech (all $p < .0001$). Across participants and stimulus types, the more detailed Step and Reduced Line models outperformed the simpler Huron and Curve models (all $p < .01$).

An optimal linear regression model was derived using stepwise deletion of predictors to minimise the Bayesian Information Criterion. This model retained musical training, stimulus type, their interaction and contour model as predictors, and accounted for 44% of variance in accuracy scores (p <.001). In addition to the effects of contour model described above, accuracy was significantly higher for melodies than for speech, and increased with musical training for melodies only. There was no interaction between musical training and contour model in this optimal regression model.

Conclusions

We explored the extent to which auditory information is summarised in the formation of mental representations of pitch sequences by comparing four melodic contour models in a novel cross-modal matching paradigm. Although the models tested were taken from the music perception literature, listeners were able to successfully match images derived from these models not only to melodies but also to spoken English sentences. The two more data-rich contour models gave rise to the highest accuracy scores, suggesting their closeness to cognitive representations. It is noteworthy that the Reduced Line model, which encodes only two sizes of pitch movements, facilitated identification no worse than the Step model, which does not compress melodic data at all.

Although participants with more musical training were more accurate in matching melodies to their visual representations, this was not the case for spoken sentences. Furthermore, musical training did not appear to affect which of the four contour models facilitated the best performance. The study supports the idea that contour plays a meaningful role in the perception and memory of music and speech, but suggests limits to the extent that musical training can bring about changes to the mental representation of pitch patterns.

Finally, the cross-modal paradigm developed here enables a rigorous comparison of formal models of contour which can be applied to any (monophonic) melody or prosodic line. Future studies could leverage the same approach to test other contour models, including those developed in the speech prosody literature.

Keywords

Contour, melody, speech, musical training.

REFERENCES

Balch, W. R., & Muscatelli, D. L. (1986). The interaction of modality condition and presentation rate in short-term contour recognition. *Perception & Psychophysics, 40*(5), 351-358.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusiclogy, 1*, 30-49.

Eerola, T. & Toiviainen, P. (2004). Mir in matlab: The midi toolbox. In *Proceedings of the 5th International Conference on Music Information Retrieval.* Available from http://ismir2004.ismir.net/proceedings/p004-page-22-paper193.pdf

Frazier, L., Carlson, K., & Clifton Jr., C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Science, 10*(6), 244-249.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience, 16*(6), 1010-1021.

Hermes, D. J. (2006). Stylization of pitch contours. In S. Sudhoff *et al.* (Eds.), *Methods in Empirical Prosody Research* (pp. 29-61). Berlin: Walter de Gruyter.

Huron, D. (1996). The melodic arch in Western folksongs. *Computing in Musicology, 10,* 3-23.

Kochanski, G. (2009). Using mimicry to learn about phonology. In E. Payne *et al.* (Eds.), *Oxford University Working Papers in Linguistics, Philology and Phonetics, Volume 12, Papers in Phonetics and Computational Linguistics* (pp.123-144). Available from http://www.ling.phil.ox.ac.uk/files/uploads/OWP2009.pdf

Mertens, P. (2004). The prosogram: semi-automatic transcription of prosody based on a tonal perception model. In B. Bel & I. Marlien (Eds*Proceedings of Speech Prosody 2004, Nara (Japan)* (pp. 549-552).

Müllensiefen, D., Gingras, B., Stewart, L. & Musil, J. (2011). The Goldsmiths Musical Sophistication Index (Gold-MSI): Technical Report and Documentation v0.9. London: Goldsmiths, University of London. URL: http://www.gold.ac.uk/music-mind-brain/gold-msi/

Müllensiefen, D., & Wiggins, G. (2011). Polynomial functions as a representation of melodic phrase contour. In A. Schneider & A. von Ruschowski (Eds.), *Systematic Musicology: Empirical and Theoretical Studies* (pp. 63-88). Frankfurt: Peter Lang

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*, 341-349.