

DATA-DRIVEN INFLUENCES ON MELODIC EXPECTANCY: CONTINUATIONS IN NORTH SAMI YOIKS RATED BY SOUTH AFRICAN TRADITIONAL HEALERS

Tuomas Eerola

Department of Music, University of Jyväskylä, Finland

ABSTRACT

This study extends a previous study concerning melodic expectations in North Sami yoiks [1] in which a comparison between expert and non-expert listeners demonstrated the existence of a core set of principles governing melodic expectancies. The previous findings are reconsidered using non-Western listeners (traditional healers from South Africa) and a set of data-driven (statistical, Gestalt and auditory) models of expectancy. This allows the roles of data-driven and schema-driven models in melodic expectancies to be separated and to reveal any possible Western bias in previous studies. The results of the experiment, in which African listeners rated the fitness of probe-tones as continuations of North Sami yoik excerpts, indicated that data-driven models are robust in explaining the expectancies, regardless of the cultural background of the listeners. Statistical models were found to have more explanatory power than the auditory model or Gestalt models.

1. INTRODUCTION

Expectancy has a central role in all forms of human behavior and it has also puzzled the minds of music theorists, psychologists and musicians. Several empirical investigations of melodic expectancies [2, 3] have demonstrated the existence of a core set of principles governing listener's melodic expectancies, captured in various models of melodic expectations [4-6]. Often these models are presumed to be universal, and this claim has been upheld in some cross-cultural studies [1, 7-9]. However, most groups of listeners in previous cross-cultural studies have been significantly exposed to Western music and hence research into the universality of the principles governing expectancies has not been completed. Also, the evidence from children suggests that although some aspects of melodic expectancies may already operate in childhood, other aspects are probably learned [10].

2. AIMS

The primary research question is the extent to which various data-driven models are similar across cultural boundaries in terms of the amount of variance in behavior they can explain. For this purpose, the study extends a previous study concerning melodic expectations in North Sami yoiks [1] using non-Western listeners (traditional healers from South Africa) and focusing on a set of core principles that have been claimed to operate despite the cultural background or learning. As this claim of innate or culture-free principles in explaining melodic expectancies is rather

controversial, this study mainly focuses on partitioning the contribution of various data-driven models relating to melodic expectancy. In this question, the explanation offered by an auditory model is especially appealing as it is based purely on automatic processing of the auditory system.

2.1 Data-Driven Models of Melodic Expectancy

The data-driven models of melodic expectancy are divided into (a) statistical models, (b) Gestalt models, and (c) auditory memory model. Each category offers a distinct way of modeling expectancies derived from the musical contexts.

Statistical models rely on the assumption that listeners adapt to the regularities of any music they hear, such as the frequency of various events in the musical context. These can be modeled with statistical variables that have been strong predictors of expectancies in previous studies [2, 9] and have also been noted to transcend cultural boundaries [11-13]. The role of statistical models in melodic expectancy should be more pronounced for the listeners who do not possess other forms of knowledge for the musical style in question. Here these factors will be modelled with two event-frequency predictors, tones and intervals in music. These predictors are not simply event counts as the durations of the events are taken into account.

Gestalt models refer mainly to the implication-realization (IR) model by Narmour [4], which has been inspired by the Gestalt-based principles of perceptual organization. The original, music-theoretic IR model describes how listeners expect upcoming tones in a melody to be proximate, similar in size, reverse the direction or fulfill other relative simple principles. The IR model originally has five factors, which have been quantified and empirically tested and extended by Krumhansl [1, 8, 9]. A simpler, *two-factor* model has been developed by Schellenberg [3], which provides equal predictive power to the *five-factor* or *seven-factor* models [3, 10, 14]. In this study the *two-factor* IR model by Schellenberg — containing two predictors, *Pitch Proximity* and *Pitch Reversal* — will be used as the most recent, simple, and successful version of the IR model.

Auditory memory model is based on the assumption that listeners use their echoic memory to ascertain whether any current musical event fits with what has previously occurred. The echoic memory has been previously incorporated by integrating new tonal events with the decayed representation of the previous tonal events [15].

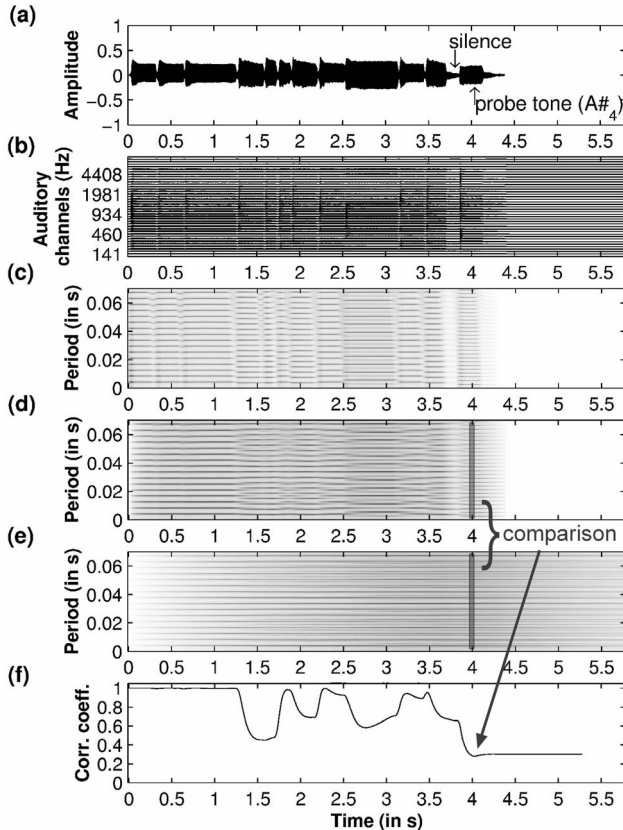


Fig 1. Pitch images for the yoik (*Anden Inga*, see top row in Figure 2) followed by a probe tone ($A\#_4$). From top to bottom: (a) audio signal, (b) auditory nerve image, (c) the periodicity pitch image, (d) the sensory memory image, (e) the echoic memory image, and (f) the comparison between sensory and echoic memory images.

Here this idea is applied by examining the fit between two types of short-term memory representations of auditory signals by the auditory system, as proposed by Leman [6]. This auditory memory model has been shown to account for the fitness ratings of the original probe-tone studies by Krumhansl & Kessler [16]. In this model, there are several stages of processing. First, the auditory periphery is simulated using an auditory model [17]. In this, the outer and the inner ear are simulated using filters that convert incoming audio signal into firing probabilities for the hair cells (*auditory nerve image*). Next, periodicity information is obtained from the auditory nerve image using an autocorrelation technique (*periodicity pitch image*). Finally, the auditory pitch image is processed using two types of short-term memories that are modelled with leaky integrators. The first memory is the *sensory memory*, extremely brief in duration and episodic in nature. The second memory is called the *echoic memory*: this includes the global auditory pitch images over the whole sequence (integrated over time). The duration of both memories (echoes) is indicated by the half-decay time of the memory (0.1 s for the sensory and 1.5 s for the echoic memory). These two memories are compared at the offset time of the probe tone by means of correlation coefficient to obtain a degree of fit of the probe-tone with the context (see Figure 1). This model extends the data-driven explanations to cover

automatic processing of auditory system, thus providing an additional explanatory level that goes beyond the pitch-based models.

2.2 Schema-driven Models

In order to evaluate the contribution of data-driven models, a simple comparison to schema-driven models will also be performed. These models are derived from the yoiks and the yoik context sequences in terms of two predictors; the first is a pentatonic predictor and the second is the correct next tone in each stimulus sequence.

3. METHOD

Participants consisted of 31 (16 men, 15 women) South African traditional Pedi healers. The participants (mean age: 42.7 years, $SD = 14$) grew up in the Pedi culture in Limpopo Province, situated in the rural, northern region of South Africa, and they have had relatively little exposure to Western music. However, all of them had been involved in musical activities (traditional Pedi music) as part of their work as religious healers. The data from the original study, where Finnish, central European, and Sami listeners rated the same excerpts, were used as a comparison data.

Anden Inga (No 1)



Elle Sunna (No 2)



Çappa Magdalena (No 3)



Haldí (No 4) 3-2 3-3



Bierra Bierra (No 5) 3-2 2-2



Fig 2. The five excerpts used in the experiment [1]. The probe positions are indicated by (PP).

Stimulus material consisted of a subset of the stimuli reported in [1], that is, North Sami yoiks (indigenous music from Northern Scandinavia). Due to practical reasons, only five out of the eight yoiks in the original study were used (shown in Figure 2). Otherwise, the stimulus material was identical to the original study (using performed excerpts that were synthesized using an English horn sample, with continuation tones consisting of the 22 chromatic scale tones in the range G_3 to E_5 , where the numeral specifies the octave; C_4 = middle C).

The procedure followed the one reported earlier [1] with minor alterations (adding graphical symbols and ensuring that the instructions are easy to translate). The task of the participants was to rate how well the continuation tone matched their expectations about what might follow in the melody, referred to below as *fitness ratings*. The procedure was carefully explained to the local ethnomusicologist, Edward Lebaka, who took care of the translation and instruction during the experiment.

4. RESULTS

The initial pool of participants numbered 31. Despite careful instructions and practice trials, the responses of six participants had to be discarded from the data analysis as outliers. The rest of the responses (25 participants) were consistent between the participants as measured by reliability analysis (*Cronbach's* $\alpha = .79$, $df = 109, 2616$, $p < .001$) and thus the mean ratings of the group will be used in further analyses of the data.

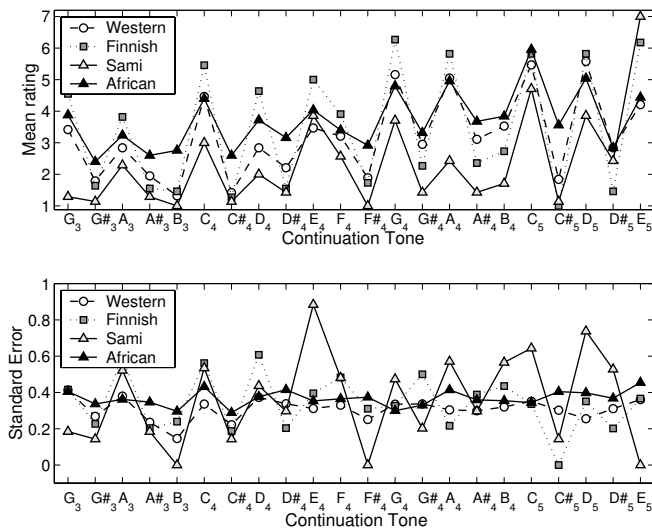


Fig 3. Yoik *Bierra Bierra* and the fitness ratings (top panel) and standard errors (lower panel) by all groups and continuation tones.

Figure 3 shows the group means for all the probe-tones for one yoik (*Bierra Bierra*), exemplifying the characteristic differences and similarities between the groups. The ratings given by the African group are most similar to those given by the Western group. The African listeners, like the Western listeners, expected the continuation tone to be very close to the last tone occurring in the context, (hence A#₄, B₄ or even C#₅), although the Western listeners did not think C#₅ would fit the musical expectancies set by the context. The lower panel displays the standard errors of ratings, indicating that the ratings of the African participants were as consistent as the responses of the other groups.

The correlation matrix of the listeners' mean ratings and predictors is investigated next (Table 1). The correlations between the participant groups demonstrate the high overall similarity of the responses. The Sami listeners deviate most from the others, as they expected certain stylistic conventions to occur in the yoiks. The greatest dissimilarity of the African ratings was with the ratings given by the Sami listeners (.71), their ratings being closer to those of the other non-experts (a correlation of .84 with the Western listeners). Both the statistical models (*Tones* and *Intervals in contexts*) and the Gestalt models (*Pitch proximity* and *Pitch reversal*) correlate better with listeners who did not possess appropriate stylistic information (Western and African) than with the other listeners. The auditory memory model, however, receives lower correlations than the statistical models and interestingly, it is the only model that fails to correlate with the Gestalt models.

Next, hierarchical regression was employed to determine whether the addition of various data-driven models improved the prediction of fitness ratings beyond that afforded by a covariate. In other words, a supplementary predictor (covariate) was entered at Step 1 of the regression to control for the extraneous variance in the responses due to the degree of consonance between the probe tone and the last tone of the context, consistent with earlier [10] investigation of melodic expectancy. In this stimulus material, consonance is an indispensable part of the musical style (in fact, a sample of 154 yoiks [18] indicates that intervals of a P4 and P5 are twice and a P8 eight times more common than the same intervals in European folksongs [$N = 6236$, [19]]). The coding of consonance was taken from Krumhansl [8] in which consonant intervals such as P8, P4, and P5 receive large values. Then the order of inclusion of the variables was determined by theoretical assumptions, taking the

	Participant ratings				Statistical models		Gestalt models	
	Sami	Finnish	Western	African	Tones in contexts	Intervals in contexts	Pitch Proximity	Pitch reversal
Sami								
Finnish	.87**							
Western	.81**	.92**						
African	.71**	.82**	.84**					
Tones in contexts	.52**	.55**	.64**	.58**				
Intervals in contexts	.40**	.42**	.48**	.46**	.42**			
Pitch proximity	.30**	.30**	.51**	.45**	.31**	.40**		
Pitch reversal	.30**	.18	.31**	.28**	.45**	.34**	.55**	
Auditory model	.27**	.34**	.34**	.28**	.38**	.19*	.02	.14

Table 1: Correlations Among the Fitness Ratings and Predictor Variables (** $p < .01$, * $p < .05$, $df = 108$ in all in cases.)

auditory model first. Table 2 displays the R^2 change values (*incremental*), that is, the predictive power of each model at each of the four steps. As means for comparison, the contribution of schema-driven models was investigated in a single, non-hierarchical regression analysis (displayed in the fifth row in Table 2).

Step and models	Sami	Finnish	Western	African
1: Consonance	.25***	.29***	.21***	.16**
2: Auditory model	.27	.34*	.26**	.20*
3: Gestalt models	.42***	.46***	.56***	.43***
4: Statistical models	.48**	.56***	.68***	.55***
Schema-driven models	.62***	.64***	.43***	.35***

Table 2. Hierarchical regression of groups of explanatory variables on fitness ratings given by the four groups of listeners (* $p < .05$, ** $p < .01$, *** $p < .001$, for R^2 change at each Step).

Table 2 indicates that the auditory model is not a particularly strong predictor of the ratings of any of the groups. The modest success of the auditory model suggests that the yoik contexts evoke a sense of fuller musical context than the cadence and scale sequences used in typical probe-tone experiments to which the auditory memory is perhaps better suited for. Gestalt models, as measured through *two-factor* IR model, seem to explain a fair amount of variance in all groups. However, the explanatory rate is highest for the Western group, implying that the model may well be biased towards Western musical structures. An alternative explanation is that the African participants imposed their own stylistic knowledge in their ratings though this cannot be modelled or tested at this stage without a comprehensive stylistic sample of the music with which they are familiar.

The statistical models were able to explain a considerable amount of variance in the regression. The contribution of this model is more pronounced in the groups who did not possess relevant stylistic knowledge. In other words, statistical information provided by the context is of real importance, especially to listeners unfamiliar with the style.

In order to appreciate the explanatory rates offered by the data-driven models, separate non-hierarchical regression results are displayed for the schema-driven (stylistic) models. In this case, the situation is the opposite of the data-driven models as the schema-driven models highlight the groups who were familiar with the musical style in question. However, the relatively high-degree of success of the schema-driven models suggests that they correlate with certain data-driven models (i.e., pentatonic predictor and tones in contexts, $r=0.47$, $p < .001$), a drawback which is common when musical excerpts are typical members of the style. This issue might nevertheless require further attention in future.

5. DISCUSSION

The results of the experiment indicate that data-driven models are robust in explaining melodic expectancies, regardless of the cultural background of the listeners. Statistical models had more explanatory power than the auditory memory model, which may be better suited to predicting the fitness ratings in key-defining contexts, in which the immediate context of chords — and not pitch patterns — induce the fitness ratings. In the present case, the yoik melodies, with their recurring musical patterns probably induce expectancies that are influenced by a longer musical context (i.e. the specific tones and intervals used in the whole excerpt) than just the match between the different auditory short-term memories.

The claim that the IR model's principles would be innate or would operate across cultures has to be approached cautiously as the current study cannot discriminate learned responses in general from innate responses, although learned expectations relating to yoik style could be distinguished. However, the success of Gestalt models in explaining Western participants' ratings especially well may be an indication of a learned heuristic. This, taken together with the evidence from other melodic expectancy and production studies from children and adults [10, 20], suggests that the Gestalt models may be viewed as learned heuristics suited to approximate musical structure. On the other hand, the contribution of statistical features in explaining melodic expectancies was larger and thus more promising than Gestalt-based heuristics. This is also in line with the findings of previous cross-cultural studies of melodic expectation [1, 9, 11-13]. It is also consistent with studies on language processing [21], in which infants as young as eight months have been shown to learn the transitional probabilities of syllables in an artificial language and tones in artificial tone sequences, during a brief exposure to them.

6. CONCLUSIONS

The present study used a cross-cultural comparison to estimate the contribution of various data-driven models to expectancy formation. An existing study was replicated with Non-western listeners, in particular South-African traditional healers, and new data-driven models were used in the analysis. An auditory memory model was only moderately successful in explaining the fitness ratings. This may be due to the nature of the stimuli, which induce true musical, sequential continuations rather than a mere comparison between the fit of the echoic memory and the probe tone. A simplified version of the IR model [3] was used to model Gestalt-based expectancies with considerable success. However, statistical models were most robust in explaining melodic expectations and future research should expand the scope of these models.

As several data-driven models were successful in explaining the expectancies of listeners from various cultures, claims of cultural relativism ([22, 23]) for musical processing are probably overemphasized. Further research on the mechanisms underlying the formation of expectations across cultures is essential in music cognition, despite the methodological and the epistemological challenges.

7. REFERENCES

1. Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition*, 76, 13-58.
2. Oram, N., & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, 57, 103-118.
3. Schellenberg, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Perception*, 14, 295-318.
4. Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago, IL: University of Chicago Press.
5. Krumhansl, C. L. (1995). Effects of musical context on similarity and expectancy. *Systematische musikwissenschaft*, 3, 211-250.
6. Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, 17(4), 481-509.
7. Unyk, A. M., & Carlsen, J. C. (1987). The influence of expectancy on melodic perception. *Psychomusicology*, 7, 3-23.
8. Krumhansl, C. L. (1995). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum*, 17, 53-90.
9. Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish Spiritual Folk Hymns: Convergence of statistical, behavioral, and computational approaches. *Music Perception*, 17, 151-196.
10. Schellenberg, E. G., Adachi, M., Purdy, K. T., & McKinnon, M. C. (2002). Expectancy in melody: tests of children and adults. *Journal of Experimental Psychology: General*, 131(4), 511-537.
11. Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, 113, 394-412.
12. Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and the West. *Music Perception*, 2, 131-165.
13. Krumhansl, C. L. (2000). Tonality induction: A statistical approach applied cross-culturally. *Music Perception*, 17(4), 461-479.
14. Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition*, 58, 75-125.
15. Huron, D., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology*, 12, 152-169.
16. Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychol Rev*, 89(4), 334-368.
17. Van Immerseel, L., & Martens, J. P. (1992). Pitch and voiced/unvoiced determination with an auditory model. *Journal of Acoustical Society of America*, 91, 3511-3526.
18. Kantola, T. (1984). Talvadaksen joikuperinne (The yoik tradition of Talvadas). In *Folkloristiikan tutkimuksia (Studies in folklore)*. Turku: University of Turku.
19. Schaffrath, H. (1995). *The Essen folksong collection in kern format. [computer database]*. Menlo Park, CA: Center for Computer Assisted Research in the Humanities.
20. von Hippel, P. (2002). Melodic-expectation rules as learned heuristics. In C. Stevens, D. Burnham, G. McPherson, E. Schubert & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney, 2002* (pp. 315-317). Adelaide: Causal Productions.
21. Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by human infants. *Psychological Science*, 9, 321-324.
22. Serafine, M. L. (1983). Cognition in music. *Cognition*, 14, 119-183.
23. Walker, R. (1996). Open peer commentary: Can we understand the music of another culture? *Psychology of Music*, 24(2), 103-130.