

For the price of a song: How pitch category learning comes at a cost to absolute frequency representations

Melody Dye, Michael Ramscar & Edward Suh

Department of Psychology, Stanford University
450 Serra Mall, Stanford, CA 94305.

Abstract

Appreciating music is cognitively demanding: listeners must learn to divide a continuous space of sound into culturally defined, discrete categories, and maintain a high degree of accuracy in their representations of those sounds. Here, we present a formal analysis of pitch category learning that reveals the trade-offs associated with learning the relative pitch categories that make music possible. Consistent with this, an empirical study reveals how under normal circumstances, people's ability to represent absolute frequency information is lost as a consequence of the learning processes that facilitate relative pitch acquisition, a finding which may help explain the rarity of absolute pitch among the general population. Understanding the contradictory computational demands of conceptual and perceptual learning can inform the design of musical training and may offer insight into the development of phonological categories in language.

Keywords: Musical Cognition, Relative Pitch, Absolute Pitch, Concepts, Categories, Learning Theory

Introduction

Most hearing people have the ability to learn musical pitch categories and the relationships between them. With training, listeners of every age can learn to distinguish types of musical intervals from one another and to identify and reproduce melodies from different starting notes, a skill known as “relative pitch” (RP). Far rarer, and virtually unknown outside the musical community, is “perfect” or “absolute pitch” (AP): the ability to accurately label a given note by its fundamental frequency (Takeuchi & Hulse, 1993). The acquisition of AP is overwhelmingly associated with musical training early in life (Chin, 2003), which has led to the suggestion that learning to label notes according to their absolute frequencies may be subject to a biological critical period (Trainor, 2005; Deutsch et al., 2006). Here, we examine whether the loss of many listeners' ability to represent absolute frequency information may be a consequence of learning to discriminate relative pitch categories, as an inherent part of the processes that make RP acquisition possible.

Musical sounds are initially encoded in the cochlea as tonotopic representations of frequency, then passed through the auditory pathway to cortex (Merzenich & Reid, 1974; Wessinger et al., 1997). Although initial sound representations are based on frequency, few listeners conceptualize music this way. Rather, most listeners make use of relative pitch (RP) systems: identifying, remembering, and producing music relative

to the differences, or intervals, in culturally defined systems of pitch.

The number of listeners able to recognize or produce tones in terms of their absolute frequencies (AF) varies by culture and task. While nearly half of music conservatory students in China can name pitches in terms of AF, the estimated rate in North American conservatory students is just 15%, and less than 0.01% for the general population (Deutsch et al., 2006). However, when tasks do not include a naming component—e.g., identifying whether familiar musical excerpts are in the correct key, or singing or humming familiar songs from memory—evidence of sensitivity to AF more widespread (Terhardt & Seewann, 1983; Halpern, 1989; Levitin, 1994; Schellenberg & Trehub, 2003). Further, speakers of tonal languages, such as Vietnamese, display a high degree of pitch consistency in their speech production across test sessions (Deutsch et al., 2004) and North Americans without musical training are even sensitive to the AF of the telephone dial tone (Smith & Schmuckler, 2008).

Interestingly, just as infants are initially able to detect and respond to the full spectrum of sounds employed in the world's languages before later losing this ability (Werker & Tees, 1984; Kuhl et al., 2006) as they acquire the sound categories of their native tongues, it appears that infants initially favor musical representations based on absolute rather than relative pitch, and that processing tends to switch over to relative as they develop (Saffran & Griepentrog, 2001). Thus, despite the popular belief that AF sensitivity in music is a special ability, mysteriously acquired by a few, it would appear that the potential for absolute pitch possession is widespread. The mystery, then, is not why so few people have it, but rather, why so many lose it (Deutsch, 2002; Miyazaki, 2004; Levitin & Rogers, 2005).

We propose that one answer lies in the nature of pitch categories, and the way they are learned. In music, melodies are usually produced and described in relation to conventionalized systems of abstract musical categories, such that melody can be appreciated independently of any particular instrument, singer or key. Chroma—musical pitch categories—do not exist in nature; rather, they are culturally determined ways of dividing up a continuous sound space. Here, we suggest that the computational demands of learning to recognize discrete structure in continuous perceptual space stands in direct opposition to the preservation of AF information.

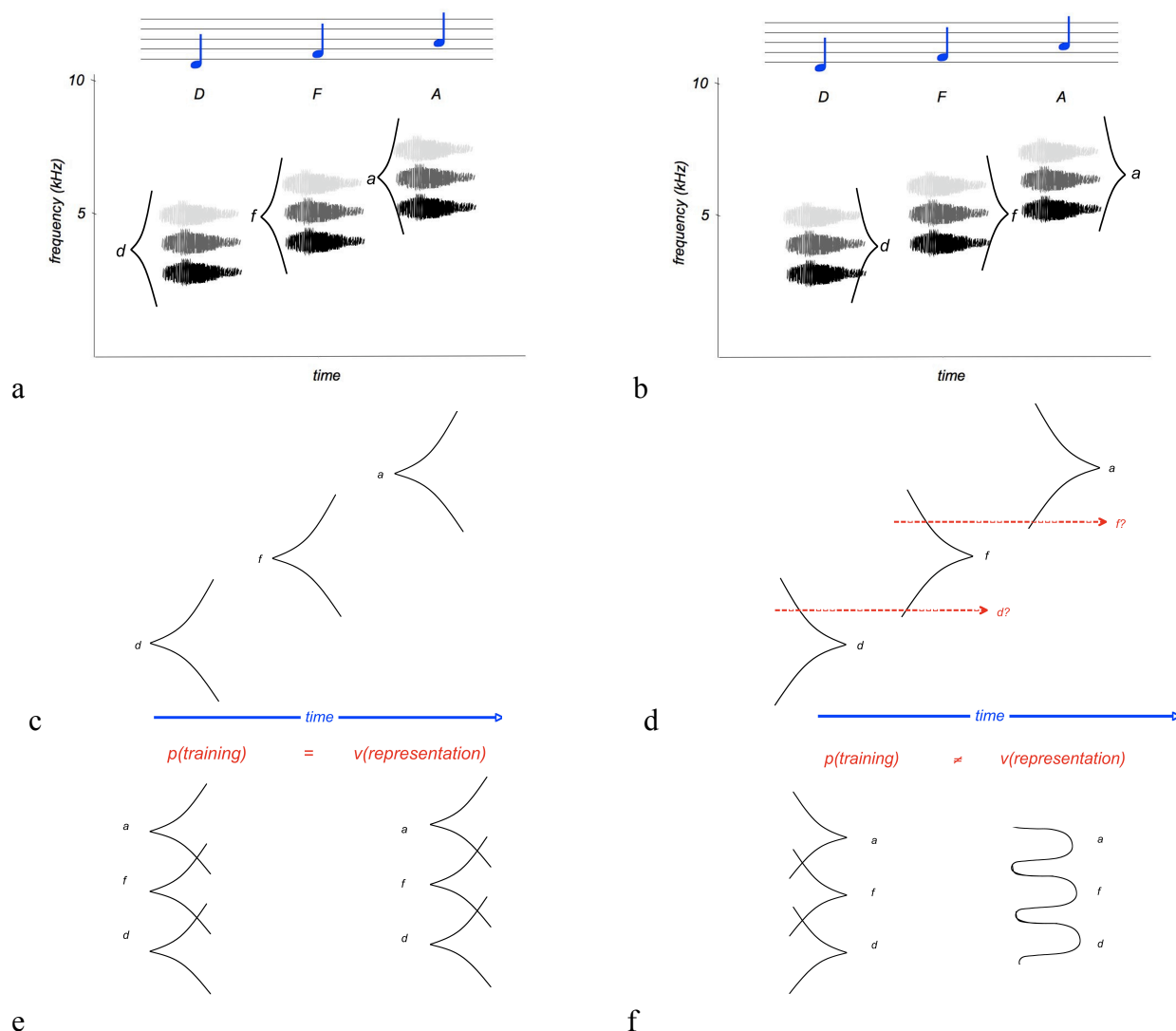


Figure 1: Labels are perceptually discrete relative to the sets of frequencies that make up complex tones (**a & b**, top panels). When tones and labels are temporally distinct (i.e., learned in sequence), two situations are possible: either the Features of a tone can predict a Label (FL) or a Label can predict the Features of a tone (LF). When a tone predicts a label (FL), the tone’s various frequencies compete as predictive cues to that label, leading to competitive discrimination learning (**d**). However, when a label predicts a tone (LF), no competition can occur, since labels are discrete cues and cannot compete with themselves (**c**). In LF-learning, the absence of competition produces a simple probabilistic representation of the signal being learned about (**e**). In FL-learning, by contrast, less reliable features in the signal lose value to those that are more reliable, resulting in representations in which some features are “overvalued” relative to their rate of occurrence, while others may be ignored (**f**). FL-learning highlights predictive features in the signal, improving pitch category discrimination while distorting absolute frequency information.

This bears elaboration. In Western music, the audible frequency continuum is divided into twelve discrete notes, spaced logarithmically. These notes repeat cyclically over the entire span of musical space. Thus exemplars of each chroma are spread across a wide frequency range (e.g., if A occurs at 220hz, it also occurs at 440hz, 880hz, etc.). Further, although there is an AF “concert pitch” convention for chromas—such that A above middle C is tuned to 440hz—in practice, orchestras tune to A over a fairly wide frequency range, from A415 in early music to A446 in some contemporary orchestras (folk and rock musicians may be even more idiosyncratic in their tuning). Importantly, regardless of the AF tuning, it is the logarithmic relationship between chroma that holds musically: i.e., if A is tuned to 440hz, D is 294hz. To

further complicate matters, musical pitch is normally conveyed in complex tones comprising both the fundamental frequency by which the pitch is identified (e.g., A440), and a range of other frequencies, which may also occur in other pitches (Hartmann, 1996).

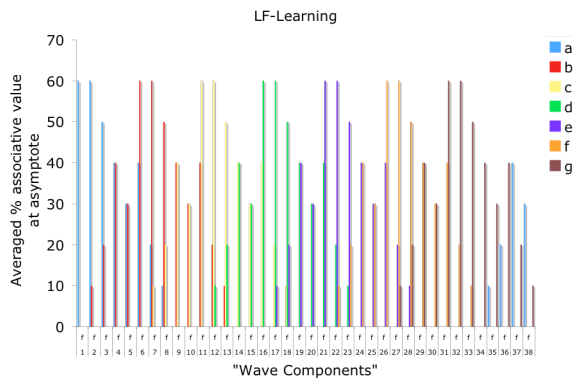
Categorizing pitch requires that the contributions of the various frequencies within a complex tone be weighted, such that its chromatic identity can be established and related to other tones (Takeuchi & Hulse, 1993). In both natural and computational learning, discriminating the more and less informative components of a complex, continuous signal for the purposes of learning a discrete category is usually achieved by adjusting the degree to which individual parts of that signal contribute to categorization (Ramscar et al., 2010). A common method for

achieving this is competitive reinforcement learning: increasing the value of parts of the signal that lead to successful predictions, and decreasing the value of parts that result in error, so that the various parts of the signal compete for learned value (Rescorla-Wagner, 1972).

The effects of this kind of competitive learning can be isolated by comparing learning from a complex stimulus to a series of discrete classes with the inverse process (Ramscar et al., 2010). As Fig.1 shows, while learning from a complex set of Features to simple Labels (FL-learning) allows for competitive learning amongst features (causing value to shift from features that produce more error to those that produce less), learning from discrete Labels to Features (LF-learning) does not allow for competition (value cannot transfer to other cues when there are none).

Computational and empirical studies have shown that these different information structures cause different representations of the relationship between features and labels to be learned (Ramscar et al., 2010). Critically, FL-learning results in better discrimination between categories because it distorts within-category representations (Nosofsky, 1991; Smith, 1989; Goldstone & Steyvers, 2001), whereas LF-learning produces poorer discrimination, but produces more veridical within-category learning (Ramscar et al., 2010). This difference offers an explanation for why people generally lose sensitivity to absolute frequency in music: AF deficits may be an inevitable part of the process of learning relative pitch (Fig.2).

LF



FL

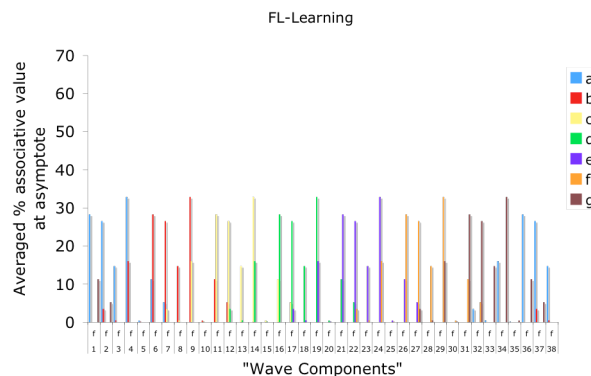


Figure 2: Simulations of the learning of identical sets of labeled categories in overlapping artificial “waveforms” structured LF (labels cue waves) or FL (waves cue labels) (Ramscar et al., 2010). The categories are labeled a-g, and the probability of a wave component occurring in each category is represented in the right panel (the LF-learned model). As can be seen, the representation of each component in the FL-learned model differs markedly from its rate of occurrence, with some components being completely unlearned relative to the labels.¹

Experiment

To examine this idea, we manipulated information structure while training undergraduates with no prior musical training to identify pitches played on a piano.

Participants

Twenty-eight Stanford undergraduates participated for course credit. Fourteen participants were randomly assigned to the Feature-to-Label training condition (Sound-first), and fourteen to the Label-to-Feature training condition (Label-first).

Training

Training consisted of a simple computer program designed to “teach” participants the names of various tones. Participants learned about the notes in a C major scale: C, D, E, F, G, A and B (Fig. A1). Notes were semi-randomly distributed throughout the training period, so that no note was ever played twice in a row (even if the note was being played in another octave).

The notes participants heard were all played on an electronic piano, and were spread over three contiguous octaves (starting at middle C, C2, and ending at C5). Because pitch-naming accuracy differs between white-key notes and black-key notes (1), only whole tones were trained. For testing purposes, each note was heard in only two of the possible three octaves. Each note was played ten times in both of the two octaves it occurred in. In total, training consisted of 140 trials.

Participants were divided between two training conditions: Feature-to-Label (Sound-first) and Label-to-Feature (Label-first). In the Feature-to-Label condition, a note was played and then its label appeared on screen (e.g., a C was sounded and then the letter C briefly

¹ In the Rescorla-Wagner (1972) model the change in associative strength between a stimulus i and a response j on trial n is defined as:

$$\Delta V_{ij}^n = \alpha_i \beta_j (\lambda_j - V_{total})$$

The model thus specifies how the associative strength (V) between a conditioned stimulus (CS_i) and an unconditioned stimulus (US_j) changes as a result of discrete training trials, where n indexes the current trial. $0 \leq \alpha_i \leq 1$ denotes the saliency of CS_i , $0 \leq \beta_j \leq 1$ denotes the learning rate of US_j , λ_j denotes the maximum amount of associative strength that US_j can support, and V_{total} is the sum of the associative strengths between all CS s present on the current trial and US_j . Learning is governed by the value of $(\lambda_j - V_{TOTAL})$ where λ_j is the value of the predicted event and V_{total} is the predictive value of a set of cues. In this simulation, all $\lambda = 100\%$, $\alpha_i = 1$ and $\beta_j = 0.2$.

appeared). In the Label-to-Feature group, the sequence was simply reversed: a label would appear on screen and then its corresponding note would be played. Each training trial lasted two seconds (Table A1). Both the Feature-to-Label and the Label-to-Feature groups were shown the exact same sequence of notes and labels. The only difference between conditions was whether the label was shown first or whether the note was played first.

Testing

Each participant was then given four tests: two tests to assess relative pitch representations and two tests to assess absolute frequency representations. Each test was comprised of 28 trials, for a total of 112 test trials. All participants completed exactly the same tests, and were given exactly the same test instructions.

Participants' **perceptual grasp of relative pitch** was tested by an octave transposition task, in which they discriminated the chroma they had been trained on from tones not heard in training. Test trials contained 7 notes that were exactly the same as those heard in training, 7 notes that were octave transpositions that had not been heard in training (but were in familiar octaves, i.e., C2-C5), and 14 lures that were half tones away from notes that had been heard within a particular octave (e.g., A4# was played rather than the A4 heard in training).

Participants' **conceptualization of relative pitch** was examined in a between-category note-name identification task: participants listened to a note being played while a label was simultaneously presented on screen, and were asked to discriminate the pairings they had seen/heard in training from novel pairings. 14 of the notes were correctly paired with their labels, 7 of the notes were critical cross-category lures (which were one note off from their label, e.g., E with F), and 7 of the notes were distractors (which were two notes off from their label, e.g., E with G). Although the note-label pairings were not always correct, participants were only tested on notes they had already heard in training (e.g., they did not hear C5 if it had not been played in training).

Participants' **perceptual discrimination of absolute frequencies** was tested in a note adjustment task. Participants listened as a note was played, and were asked to discriminate the notes they had heard in training from notes that they had not been heard in training. In this test, 14 of the notes were those that had been heard in training and 14 of the notes were "adjustments" of correct notes heard in training (the new notes were all slightly sharp, see below).

Participants' **conceptual representations of absolute frequencies** were tested by means of a within-category note-name discrimination task: as in the between-category task, participants heard a note played while a label was being presented simultaneously on screen, and were asked to discriminate pairings they

had seen/heard in training from new pairings. However, in this test, the critical lures were slightly "adjusted" versions of the original note paired with the original labels (e.g., a sharp B with the label B). Participants had to discriminate 7 of these within-category adjusted notes from 7 original notes that were correctly paired with their labels (e.g., a perfect B with the label B). To make this a particularly rigorous examination of our participants' representations of absolute frequencies, there were an additional 14 distractors in this test: 7 of which were "adjusted" notes paired with the the note they had been moved towards (e.g., a slightly sharp B with the label C), and 7 of which were "adjusted" notes paired with the note they had been moved away from (e.g., a slightly sharp B with the label A). This meant that 75% of the test items were in tune with one another, while being out of tune with the training items. All the notes played were either notes played in training, or adjustments of notes played in training, and all the labels had been seen in training (i.e. no octave transpositions were tested).

Results

For analysis purposes, correct and false-positive response rates were used to estimate recognition sensitivity (d' ; (MacMillan & Creelman, 1991) Fig.3). A 2 (training condition) \times 2 (absolute / relative frequency test) analysis of these rates revealed a main effect of task (participants performed better on absolute frequency tests; $F(1,26)=5.068$, $p<0.05$) and an interaction between training and testing ($F(1,26)=9.593$, $p<0.0005$): FL-trained participants performed better on relative frequency tests ($t(26)=1.859$, $p<0.05$), whereas LF-trained participants performed better on absolute frequency tests ($t(26)=2.212$, $p<0.05$).

Analysis of the perceptual tests revealed a straightforward interaction between performance and training ($F(1,26)=18.272$, $p<0.0001$; see Fig.4), whereas, this interaction was not significant in the conceptual tests ($F(1,26)=1.961$, $p>1.7$). Instead, there was a main effect of training ($F(1,26)=6.701$, $p<0.05$), confirming that in the conceptual tests, FL-training raised participants' relative performance to above chance levels ($t(13)=2.546$, $p<0.05$; LF-trained, $t(13)=1.231$, $p>0.2$) at the expense of degraded absolute performance in the absolute conceptual test, where FL-trained participants performed below chance ($t(13)=-2.546$, $p<0.05$; LF-trained $t(13)=0.072$, $p>0.9$; Fig.4). While LF-trained participants were unable to discriminate the correct notes, their performance is still notable: this test was strongly biased against absolute responding (75% of the items tested were sharp lures or distractors that were in tune with each other, but out of tune with the correct notes), and this relative bias caused the FL-trained participants to misidentify the sharpened notes far more often than they identified the correct notes ($t(13)=2.59$, $p<0.05$).

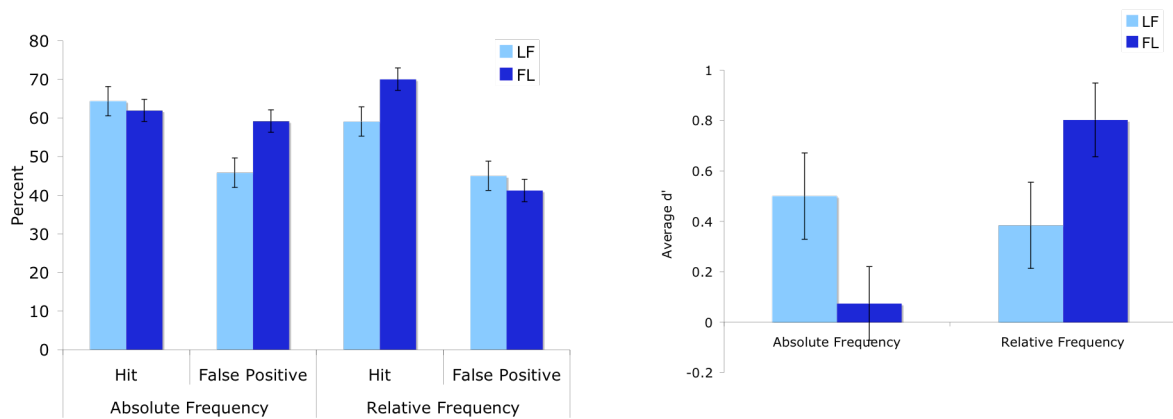


Figure 3: Average rates of correct and false-positive responses in the tests (right panel) were used to calculate d' estimates of recognition sensitivity (left panel): LF-trained participants performed better on the absolute frequency tasks, whereas FL-trained participants performed better on the relative pitch tasks (error bars are SEM).

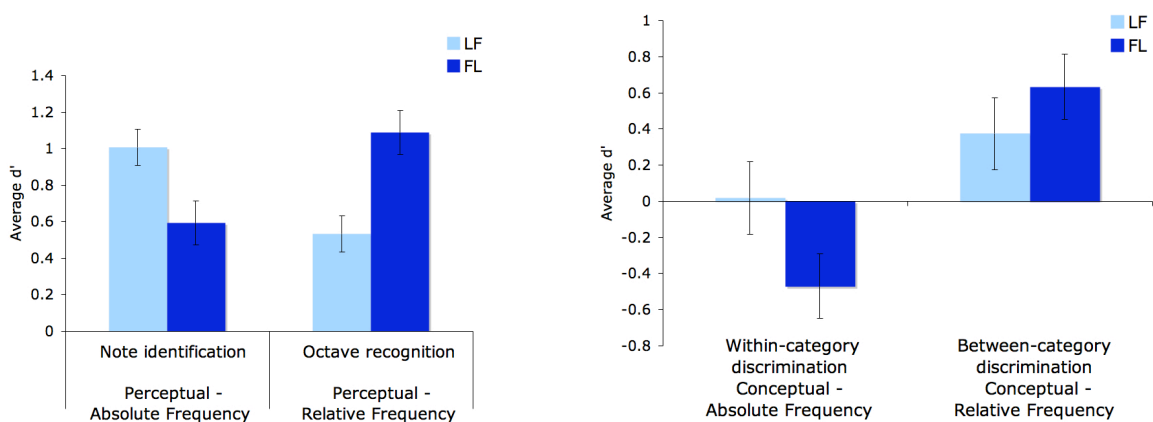


Figure 4: The cause and consequence of a trade off: LF and FL-trained participants on the two perceptual discrimination tasks (left panel), and the conceptual tasks (right panel). In the perceptual test, LF-trained participants better distinguished concert pitch notes from sharp lures, whereas FL-trained participants better discriminated correct notes—and octave transpositions of correct notes—from lures. The change to the underlying representations that underlie this trade off can be seen in the results of the conceptual test, where the improvement in between-category discrimination performance brought about by FL-training came at a cost to within-category discrimination (error bars are SEM).

Discussion

As has often been noted, the mystery about absolute pitch (AP) is not why so few have it: it's why most don't. These results provide an explanation: there is a price to pay for learning relative pitch categories—the loss of absolute frequency sensitivity (Miyazaki, 2004; Levitin & Rogers, 2005). Learning pitch categories caused our FL-trained participants to ignore—and lose much of their ability to discriminate—the absolute frequencies of the notes that they heard.

That our participants learned the perceptual categories far better than the conceptual categories can also help explain why AP is usually acquired early in life. Our results suggest that learning the conceptual categories may depend on first learning the right perceptual discriminations. Given that learning to perceptually discriminate pitch categories impairs people's ability to represent absolute frequency, it follows that when perceptual and conceptual learning are *decoupled*—as happens commonly when people are

exposed to music without first learning the names of notes—subsequent learning will be based on representations that contain less absolute frequency information.

At the same time, exposure to the sounds of musical instruments prior to (or simultaneous with) learning about music itself may result in the development of representations of those sounds alongside the learning of musical categories. This idea is consistent with the overwhelming tendency for AP to be associated with early musical training, with the finding that many AP possessors are limited to AP on the instrument they were trained on as children (Miyazaki, 2004), and with a range of findings on AP naming, revealing that while AP listeners are accurate at identifying pitch (subject to influences of timbre and pitch range), they have difficulty perceiving pitch relations in different pitch contexts and in recognizing transposed melodies, as compared to listeners without AP (Miyazaki, 2004). In the same way that learning leads to a loss of absolute

frequency representations to listeners without AP, our analysis suggests that preserving absolute frequency in musical representations ought to impair relative discrimination and processing. This highlights a simple principle: there can be ‘no representation without taxation’ in the development of musical pitch representations (see also, Miyazaki, 2004).

While our results certainly do not rule out a maturational or genetic component in AP ability (Zatorre, 2003; Theusch, Basu & Gitschier, 2009), they underline the importance of understanding the conflicting demands of discrimination and representational fidelity in learning, and the implications this has for our understanding of representation. Here, we have shown how principles of information and learning can be used to illuminate some of the puzzles of pitch perception. The systematic application of these principles to other problems in cognition may shed light on a much deeper mystery: what it is that our human capacity to “represent our environment” actually involves.

Acknowledgements

This material is based upon work supported by NSF Grant Nos. 0547775 and 0624345 to Michael Ramscar.

References

- A.H. Takeuchi, S.H. Hulse, Absolute pitch, *Psychol Bull* 113, 345-361, (1993).
- A. R. Halpern, Memory for the absolute pitch of familiar songs, *Mem. Cognit.* 17, 572-81 (1989).
- C.M. Wessinger, M.H. Buonocore, C.L. Kussmaul, G.R. Mangun, Tonotopy in human auditory cortex examined with functional magnetic resonance imaging, *Hum. Brain Mapp* 5, 18-25 (1997).
- C.S. Chin, The development of absolute pitch: A theory concerning the roles of music training at an early developmental age and individual cognitive style, *Psychology of Music* 31, 155-171 (2003).
- D. Deutsch, T. Henthorn, E. Marvin, H. Xu, Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period, *J. Acoust. Soc. Am* 119, 719-22 (2006).
- D. Deutsch, T. Henthorn, M. Dolson, Absolute pitch, speech, and tone language: Some experiments and a proposed framework, *Music Perception* 21, 339-56 (2004).
- D. Deutsch, The puzzle of absolute pitch, *Curr. Dir. Psychol. Sci.* 11, 200-4 (2002).
- D. J. Levitin, Absolute memory for musical pitch: evidence from the production of learned melodies, *Percept. & Psychophys.* 56, 414-23 (1994).
- D. J. Levitin, S.E. Rogers, Absolute pitch: perception, coding, and controversies, *Trends Cogn Sci.* 9, 26-33 (2005).
- E. G. Schellenberg, S. E. Trehub, Good pitch memory is widespread, *Psychol. Sci.* 14, 262-6 (2003).
- E. Terhardt, M. Seewann, Aural key identification and its relationship to absolute pitch, *Music Percept.* 1, 63-83 (1983).
- E.V. Theusch, A. Basu, J. Gitschier, Genome-wide study of families with absolute pitch reveals linkage to 8q24.21 and locus heterogeneity, *Am J Hum Genet* 85, 112-19 (2009).
- Flemming, E. Contrast and perceptual distinctiveness. In Hayes, Kirchner & Steriade eds., *Phonetically Based Phonology*. (Cambridge University Press, Cambridge, 2004).
- J. F. Werker, R. C. Tees, Cross-language speech perception: Evidence for perceptual reorganization during the first year of life, *Infant Behav. Develop.* 7, 49-63 (1984).
- J.R. Saffran, G.J. Griepentrog, Absolute pitch in infant auditory learning: evidence for developmental available reorganization, *Dev Psychol.* 37, 74-85 (2001).
- K. Miyazaki, How well do we understand absolute pitch?, *Acoust. Sci. & Tech.* 25, 426-32 (2004).
- L. B. Smith, *Psych. Rev.* 96, 125 (1989).
- L.J. Trainor, Are There Critical Periods for Musical Development?, *Devel. Psychobiol* 46, 262-78 (2005).
- M.M. Merzenich, M.D. Reid, Representations of the cochlea within the inferior colliculus of the cat, *Brain Res* 77, 397-415 (1974).
- M. Ramscar et al., The Effects of Feature-Label-Order and their Implications for Symbolic Learning, *Cognitive Science* (2010). DOI: 10.1111/j.1551-6709.2009.01092.x
- N. A. Macmillan, C. D. Creelman, *Detection theory: A user's guide* (Cambridge University Press, Cambridge, 1991).
- N. A. Smith, M.A. Schmuckler, Dial A440 for absolute pitch: Absolute pitch memory by non-absolute pitch possessors, *J. Acoust. Soc. Am.* 123, (2008).
- P.K. Kuhl et al., Infants show a facilitation effect for native language phonetic perception between 6 and 12 months, *Develop. Sci.* 9, F13-F21 (2006).
- R.A. Rescorla, A.R. Wagner in *Classical Conditioning II: Current Research and Theory*, Black & Prokasy Eds., (Appleton-Century-Crofts, 1972).
- R.J. Zatorre, Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function, *Nat. Neurosci.* 6, 692-5 (2003).
- R.L. Goldstone, M. Steyvers, The Sensitization and Differentiation of Dimensions During Category Learning, *JEP: Gen.* 130, 116-39 (2001).
- R. M. Nosofsky, Stimulus bias, asymmetric similarity, and classification, *Cognitive Psychology* 23, 94-140 (1991).
- W.M. Hartmann, Pitch, periodicity, and auditory organization, *J. Acoust. Soc. Am.* 100, 3491-3502 (1996).