DIATONIC CATEGORIZATION IN THE PERCEPTION OF MELODIES

Jason Yust,

University of Alabama

ABSTRACT

Previous research has shown that musicians skilled at interval identification categorize intervals according to their approximate size in semitones to perform interval discrimination tasks. The hypothesis that listeners have a similar strategy, diatonic categorization (a categorization scheme with seven rather than twelve categories within the octave), available to them for a range of other musical tasks neatly explains some phenomena of musical practice and some previous experimental results. Diatonic categorization would be a particularly effective strategy for music based on diatonic scales, and because it is a more efficient representation, it might be useful to listeners with less musical training and for musical tasks that are more complex than simple interval identification. I tested the specific hypothesis that musicians with moderate training would use diatonic categorization as a strategy in a short term memory task for transposed "melodies" (tone sequences) using two types of stimuli, one with diatonic stimuli in a familiar twelve-tone equal tuning (12-tET) and one with mock-diatonic stimuli in an unfamiliar thirteen-tone equal tuning (13-tET). The results show that subjects used diatonic categorization in the unfamiliar tuning condition as the only effective strategy to recall the interval content of melodies. In the familiar tuning, diatonic categorization appeared to be accompanied by other strategies, though more intersubject variability made these results less clear.

1. BACKGROUND AND AIMS

Auditory categorical perception was first observed by Liberman et al. (1957) in the perception of speech consonants. Since then categorical perception has been observed in numerous aspects of speech perception and the concept has been greatly refined and modified. (Lindblöm 1996, Macmillan 1987, Rosen & Howell 1987) The basic elements of categorical perception research are identification and discrimination paradigms. Where there is categorical perception along some physical continuum, subjects (1) are able to use categories to *identify* where stimuli occur on the continuum, and (2) are able to *discriminate* stimuli that fall into different categories, but not ones that fall within the same category.

Burns & Ward (1978) found an instance of categorical perception in a specifically musical context, the task of interval identification, and other studies have confirmed their basic findings (Siegel & Siegel 1977, Burns 1999). The categories in this case are the twelve semitonal interval sizes between unison and octave. In identification, subjects classify intervals presented in sinusoidal tones in familiar musical categories ("minor second," "major third," etc.). In discrimination, subjects hear pairs of intervals presented at different pitch levels and classify them as same or different. It is important in both cases that the intervals are transposable. Even though the discrimination task should be simpler from a signal theory perspective, musicians are better at the more familiar task of identification (Burns & Campbell 1994), suggesting strong categorical perception: i.e., discrimination can only be performed on the basis of identification. Yet the phenomenon is only observed for very skilled musicians who have experience in interval identification. The untrained subjects in Burns & Ward 1978 showed no evidence of categorical perception in addition to performing poorly on the task. Therefore, while twelve-tone categorization may be the only strategy available for the precise discrimination of isolated intervals, other strategies must exist for the discrimination of melodies, since presumably both musicians and non-musicians can perform such tasks to some extent.

The hypothesis motivating the present study is that a different type of interval categorization strategy, using only seven interval categories per octave, helps listeners recall the relative pitch characteristics of melodies. I will call this strategy *diatonic categorization*. Because diatonic categorization reduces the number of categories, it is a more efficient representation for melodies that demonstrate a relatively high degree of conformity to diatonic scales (or other relatively even seven-note scales). Because of the reduced cognitive load of the diatonic representation, this type of categorization may be available to less trained listeners who cannot always accurately identify the semitone categories of intervals.

There is a great deal of circumstantial evidence in support of the diatonic categorization hypothesis. Most obviously, it is intrinsic to Western interval nomenclature, where the generic name for each semitonal interval size includes its diatonic category ("second," "third," et c.—the exception being "tritone" which is the generic name for an interval that falls on a diatonic category boundary).

Empirical research provides additional circumstantial evidence. For instance, in interval identification experiments subjects more often confuse diatonic-category equivalents than other intervals differing in width by a semitone. (Plomp, Wagenaar, & Mimpin, 1973; Killam, Lorton, & Schubert, 1975) In a series of experiments, Balzano (1977) found that subjects could identify intervals more quickly when using generic (e.g., "second") rather than specific (e.g., "major second") interval names. If they were identifying intervals in semitonal categories, than the use of generic names should have required an extra cognitive step and therefore should have taken longer. (See also Balzano, 1982a–b). Finally, numerous studies on transposed-melody discrimination demonstrate that melodies that differ only in mode are particularly difficult to discriminate, especially for musically untrained listeners. (Dowling and Bartlett 1981, Dewitt & Crowder 1986, Halpern 1984, Halpern, Bartlett, & Dowling 1998, and Leaver & Halpern 2003) Changes in mode preserve the diatonic categories of all intervals in a melody.

The present study looks for evidence of diatonic categorization in a melody discrimination task. Melody discrimination should be more complex as a short-term memory task than interval identification but it also more closely resembles familiar musical tasks.

2. METHOD

Subjects were 36 undergraduates of the University of Washington enrolled in a first-year theory course for music majors. All subjects had a substantial musical background and knowledge of the rudiments of music theory. However, they offered a wide range and variety of musical abilities and were not screened for any particular ability (such as interval identification skills). Subjects participated voluntarily and received ear-training credit for their participation. All subjects were questioned before the test as to whether they possessed absolute pitch abilities, and the results of two subjects who claimed substantial absolute pitch skills were left out of the data pool. Furthermore, data was not obtained from two subjects due to failure to follow instructions.

Each subject took the test binaurally at a Macintosh desktop computer with headphones. The stimuli were synthesized and the test conducted using SuperCollider. The session began with instructions and a practice session with feedback. Subjects were allowed to adjust the volume during the practice session. They made responses and initiated trials using a mouse. The stimuli consisted of 500 ms enveloped sinusoidal tones. Each "melody" included six tones. A trial consisted of a standard melody and a comparison melody, separated by a one second pause. The subjects were informed that "different" conditions would always involve a change in the final note of the comparison melody.

There were two tuning conditions. In the "familiar tuning" condition stimuli conformed to diatonic scales in 12-tone equal tuning. The "unfamiliar tuning" condition instead used a mock-diatonic scale in 13-tone equal tuning. This scale, with step pattern 2-2-2-2-2-2-1, is the maximally even 7-note scale of 13-tET (see Clough & Douthett 1991). In addition to being relatively even, it shares another important property with the diatonic: just as one can move between diatonic scales along the circle of fifths by changing one note by one semitone (such as going from C major to G major by changing F to F♯), one can move between different versions of this 13-tET scale by moving the note on either side of the 13-tET semitone in the scale. The 13-tET scale should facilitate diatonic categorization in the same way as the diatonic—the diatonic category of each interval corresponds to its scale-step span (i.e., how many scale-steps it encompasses). However, it should confound attempts at twelve-tone categorization, since intervals will not add in a way that is consistent with their twelve-tone categories, and the intervals from five to eight thirteenth-octaves are close to the category boundaries for semitonal categorization.

The use of sinusoidal tones was primarily intended to eliminate the distracting "out-of-tune" character that this 13-tET tuning would elicit if complex tones were used, since sinusoidal tones eliminate the possibility of hearing misaligned upper partials. Indeed, all subjects were questioned about the tuning of stimuli after the experiment and only one showed any awareness of anything unusual in the tuning, or of the fact that different tunings were used.

In addition to the scale condition, stimuli were systematically constructed according to two contour conditions and eight possible final intervals. Here are the stimuli as they would appear in C major (in the familiar tuning condition):

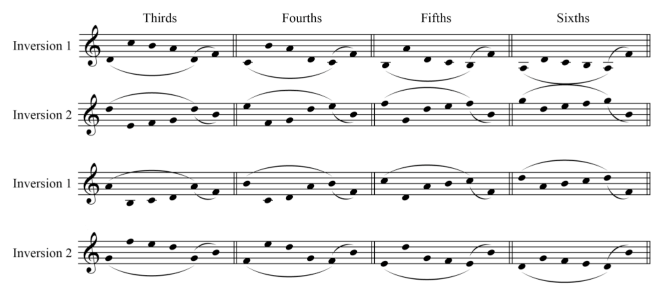

Figure 1: Stimuli as they would appear in C major.

The ending note is $\hat{7}$ (B) or $\hat{4}$ (F), so that when it is changed by a semitone in one direction, it is a different note in the same scale, and when it is changed by a semitone in the other direction, it changes the scale to G major or F major. In the former case, the diatonic categorization of every interval involving the last note changes. I will call this the DCAT condition, meaning that diatonic categorization is a potentially effective strategy. The latter case is the "NO-DCAT" condition because diatonic categories do not change for any intervals. The first five notes include all the notes of the scale other than the ones a step away from the final. The first note is repeated as the penultimate note, which serves to emphasize one particular interval. Since there are four possibilities and they can occur either as ascending or descending intervals, in addition to the contour and tuning factors, there are a total of 32 melodies. Each subject would hear each of these three times—a same trial, a DCAT trial, and a NO-DCAT trial—for a total of 96 trials. The 13-tET stimuli were constructed in the same way, with the final notes chosen so that moving it up or down by a 13-tET semitone would create the same DCAT/NO-DCAT distinction. Because there are two such possible notes in both scale types and because of the balance of ascending and descending final intervals, the DCAT/NO-DCAT conditions do not correlate with the direction of change in either the final tone or the size of the final interval.

3. RESULTS

The d' statistic gives the most accurate account of performance for perceptual discrimination tasks. The same-different paradigm used in the present experiment best fits the "differencing strategy" described by Macmillan (2002). The "differencing" model assumes that the listener observes standard and comparison independently and then evaluates their psychological distance according to some criterion.

A 2 × 2 repeated measures ANOVA comparing d' scores on the diatonic categorization conditions and tuning conditions revealed a significant effect of diatonic categorization ($F = 6.097$, $\eta = .41$, $p < .05$), and marginal effects for tuning ($F = 3.543$, $\eta = .32$, $p = .069$) and tuning × DCAT ($F = 3.543$, $\eta = .33$, $p = .061$). Figure 2 illustrates these effects. The effect of DCAT was in the predicted direction, as was the marginal effect of tuning, but the marginal interaction suggests that the effect of DCAT was primarily due to the 13-tET condition. Indeed, a simple effects analysis of DCAT by tuning reveals a large significant effect of DCAT in the 13-tET condition ($F = 10.375$, $\eta = .50$, $p < .005$), and no significant effect in the 12-tET condition ($F = 0.903$). As Figure 2 illustrates, performance was significantly above chance in all conditions except the 13-tET condition without diatonic categorization. Table 1 gives the means calculated for each d'.

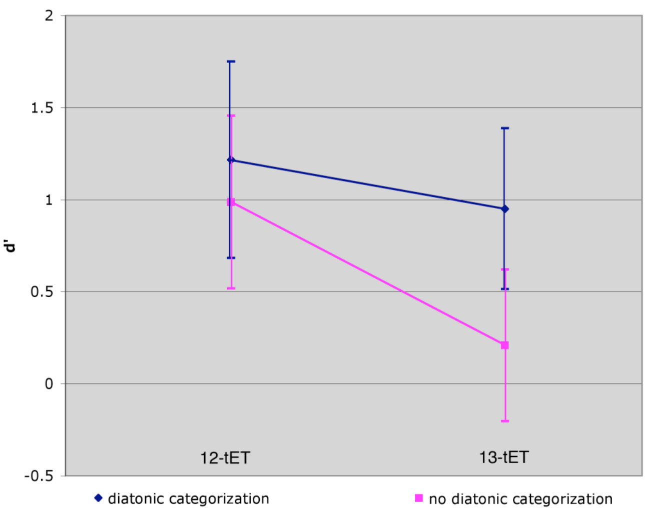

Figure 2: Mean d's and confidence intervals for each tuning × DCAT condition.

Condition	d'	SD	T	Sig.
12tET, DCAT	1.22	1.48	4.65	($p < .001$)
12tET, NO-DCAT	0.99	1.30	4.29	($p < .001$)
13tET, DCAT	0.95	1.21	4.44	($p < .001$)
13tET, NO-DCAT	0.21	1.14	1.04	—

Table 1: Mean d's and significance for each tuning × DCAT condition.

The repeated measures ANOVA also revealed a great deal of inter-subject variability. Although there was no overall between-subjects effect in the d' scores ($F = 0.763$), there were significant interactions between subjects and tuning ($F = 4.465$, partial $\eta = .90$, $p < .001$) and between subjects and DCAT ($F = 2.252$, partial $\eta = .83$, $p < .05$). This suggests that subjects varied widely in their listening strategies.

A 2 × 2 (tuning × DCAT) ANOVA blocked by subject on the response times of every correct response yielded equivocal results. Neither the effect of DCAT ($F(2, 65) = 0.593$) or tuning ($F(1, 32) = 1.352$) reached significance, and while the mean times on DCAT followed the predicted order of NO-DCAT > DCAT > "same" (with the mean difference between NO-DCAT and DCAT at a mere 16 ms and the DCAT/"same" difference at 109 ms), the difference in mean times on tuning went in the opposite direction, with response times on the familiar tuning exceeding those on the unfamiliar tuning by 120 ms on average.

The analysis of response times on correct responses reinforced the observation of inter-subject variability in performance. Unsurprisingly, there was a large overall between-subjects effect ($F(31,43) = 2.203$, partial $\eta = .78$, $p < .01$). More interestingly, the interactions of subject × tuning ($F(31, 65) = 2.056$, partial $\eta = .70$, $p < .01$) and subject × DCAT ($F(62, 62) = 1.988$, partial $\eta = .82$, $p < .01$) showed significant effects in the response time data as in the performance data.

Planned polynomial trend contrasts on the size of the final interval showed significant linear ($F = 10.188$, $\eta = .50$, $p < .005$) and quadratic ($F = 5.476$, $\eta = .39$, $p < .05$) trends in the 12-tET condition, and a marginal quintic trend ($F = 3.652$, $\eta = .32$, $p = .065$). See Figure 3. The linear trend reflects the expected decreasing performance on larger intervals. The quadratic trend and the marginal quintic trends resulted from relatively low performance on the major third and perfect fifth standards and the relatively high performance on the minor sixth and minor third relative to the linear trend. The 13-tET condition gave no significant main effect for size of final interval in the one-way repeated measures ANOVA.

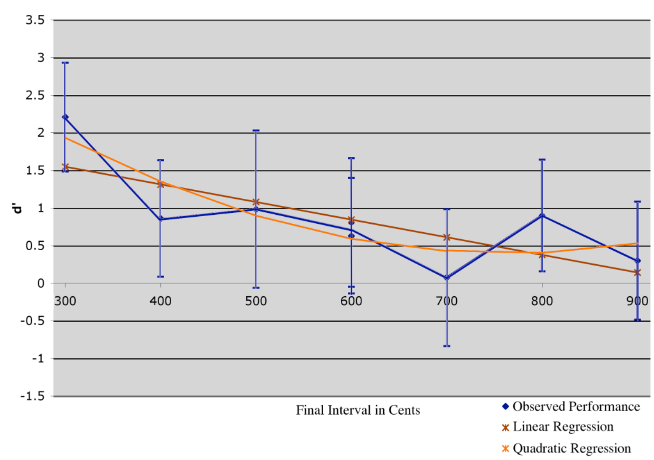

Figure 3: Mean d's and confidence intervals for the size of the final interval in the 12-tET condition.

Since the main analysis indicated a possible difference between tuning conditions, I did post-hoc tests on whether twelve-tone categorization played a role in performance by comparing the tuning conditions on two types of intervals: those with final intervals in 13-tET that fall into unambiguous twelve-tone categories (thirds and sixths) and those with ones that are ambiguous (fourths and fifths). The analysis yielded the opposite of the expected result: there was no significant difference between tuning conditions for fourths and fifths ($M = 0.01$, $SD = 1.98$, $t = 0.04$) and a significant (at a Bonferroni adjusted α) advantage for 12-tET on thirds and sixths ($M = 0.82$, $SD = 1.79$, $t = 2.591$, $p < .025$). See Figure 4.

I then performed a 2 × 2 × 2 (interval type × DCAT × direction of alteration) ANOVA on response times blocked by subject, which revealed a three-way interaction ($F(1, 25) = 4.729$, partial $\eta = .40$, $p < .05$) reflecting shorter reaction times on third/sixth melodies where the standard ends on $\hat{7}$ than where it ends on $\hat{4}$, and the opposite tendency for fourth/fifth melodies. Performance on 12-tET third/sixth melodies, where particularly strong performance (average *d'* of 1.73) was observed for diatonic categorization with a standard ending on $\hat{7}$, is consistent with this reaction time data, but the performance on 12-tET fourth/fifth melodies generally trended opposite the reaction time data.

4. CONCLUSIONS

The results confirm the diatonic categorization hypothesis, showing that subjects used diatonic categorization as an effective strategy, particularly in the unfamiliar tuning condition. The overall effect of diatonic categorization accounted for a moderate proportion of variance, but when broken down by tuning condition we find a small non-significant effect in the familiar tuning and a large effect in the unfamiliar tuning. Furthermore, performance was at chance in the unfamiliar tuning without diatonic categorization, suggesting that diatonic categorization was the only effective strategy available to listeners in the unfamiliar tuning, whereas strategies specific to twelve-tone equal tuning obscured the effect of diatonic categorization in the familiar-tuning context. Post-hoc tests attempted to tease out some of the other strategies specific to the familiar tuning.

This is a strong and perhaps surprising result in light of previous research and theories of relative pitch. It is commonly asserted that listeners use a diatonic schema to represent stimuli that conform to a diatonic scale—see, e.g., Jordan & Shepard (1987). Dowling (1991) and Dowling et al. (1995) argue, for instance, that listeners encode tonal melodies in terms of a succession of scale-degree numbers rather than in terms of intervals. This assertion is based largely on results involving recall of melodies over long filled delays and having a relatively high degree of "tonal structure," and therefore may not necessarily be applicable to the present experiment. However, even over short delays, alterations to transposed melodies that violate diatonic scalarity are more easily perceived than those that do not, for both relatively structured and relatively unstructured diatonic tone sequences (Cuddy, Cohen, & Miller, 1979).

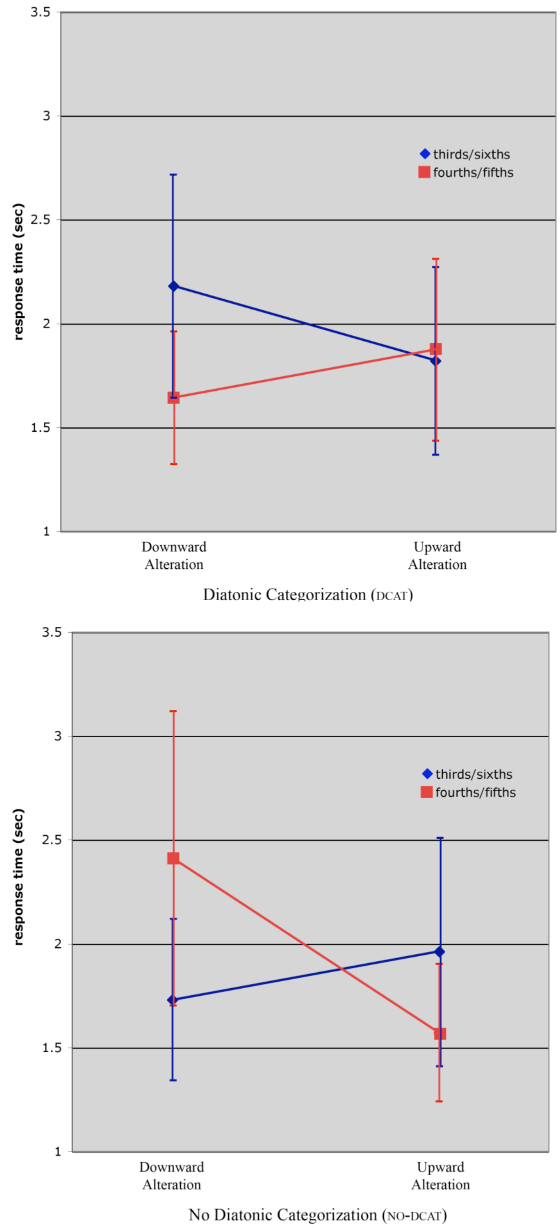

Figure 4: Response times and confidence intervals for correct responses in 12tET conditions, separated by interval type (thirds/sixths vs. fourths/fifths) × DCAT × direction of alteration.

All of this might lead one to predict a result opposite of the one observed in the present experiment. If subjects had used some sort of diatonic schema to encode standard tone sequences, the final note in the NO-DCAT condition would have sounded like an out-of-key note, whereas in the DCAT condition the final note would have been in-key. However, subjects found it easier to hear the in-key alteration, suggesting that they did not use comparison to a diatonic schema as a strategy.

In the 12-tET condition performance was significant in both categorization conditions, indicating that there were listening strategies other than diatonic categorization available in the 12-tET condition not available in the 13-tET condition. Many subjects reported attempting to use an interval-identification strategy (against the recommendation of pre-test instructions). This strategy would be effective in the 12-tET condition for trained musicians, but in the 13-tET condition subjects would have been confused by the extremely out-of-tune perfect fourths and fifths. As Siegel & Siegel (1977) have shown, skilled interval identifiers are generally unable to distinguish sharp perfect fourths from flat perfect fourths, and likewise perfect fifths. If twelve-tone categorization of the final interval were an important strategy, we should find a stronger difference between tuning conditions in the trials with fourths or fifths (rather than thirds or sixths) as final intervals.

Yet, the results fail to support this hypothesis; performance in the familiar tuning was significantly better for sequences with a third or sixth as a final interval and was indistinguishable from performance on the unfamiliar tuning stimuli when the final interval was a fourth or fifth. The explanation for this might be that 12-tET stimuli facilitated a “modal feeling” listening strategy not quite comparable to simple semitonal categorization. Indeed, multiple subjects reported listening for “majorness” or “minorness” after the experiment. The contours with thirds or sixths as final intervals would have given a stronger modal impression (because the tend to parse into triads involving the repeated note and final note). Post hoc examination of reaction times in the familiar tuning condition corroborated the “modal hearing” hypothesis. The third/sixth melodies ending on $\hat{7}$ suggest dominant harmony and those ending on $\hat{4}$ suggest the less familiar supertonic harmony, while the fourth/fifth melodies do not generally evoke strong tonal implications. The shorter reaction times and better performance on 12-tET third/sixth melodies where the standard ends on $\hat{7}$ could reflect the influence of a perceptual asymmetry related to familiarity of harmony (c.f. McFadden & Callaway 1999).

The finding of strong inter-subject variability in the effects of both tuning and diatonic categorization on performance suggest that subjects differed consistently in their listening strategies. The subject/tuning and subject/DCAT interactions in reaction times reinforce this conclusion.

In summary, then, the “mistuning” of stimuli in 13-tET was an effective experimental strategy for interfering with tonality-specific melody discrimination strategies and revealing the effects of a diatonic categorization. The 12-tET stimuli, on the other hand, admitted strategies other than diatonic categorization. These strategies are apparently distinct from the basic twelve-tone categorization observed by Burns & Ward (1978) in simple interval identification and discrimination paradigms, though the exact nature of the strategies is not entirely clear from the present data.

5. ACKNOWLEDGMENTS

I would like to thank Steven Demorest for all of his help in constructing and perfecting the paradigm for this experiment, and Joshua Parmenter for his immeasurable genius for writing Supercollider code. Thanks also to John Miamoto for advice on statistical analysis, to Tom Baker for making it possible to recruit subjects, and to all the students of Music 203 at the University of Washington for their participation and interest in the study.

6. REFERENCES

Balzano, G. J. (1982a) Musical vs. psychoacoustical variables and their influence on the perception of musical intervals. *Bulletin of the Council for research in music education 70*, 1–11.

——— (1982b) The pitch set as a level of description for studying musical pitch perception. In M. Clynes (Ed.), *Music, Mind, and Brain: The Neurophysiology of Music* (321–351). New York: Plenum.

——— (1977) Chronometric studies of the musical interval sense. PhD Dissertation, Stanford University.

Burns, E. M. (1999) Intervals, scales, and tunings. In D. Deutsch (Ed.), *The Psychology of Music* (215–264). San Diego: Academic Press.

Burns, E. M. & Ward, W. D. (1978) Categorical perception--phenomenon or epiphenomenon?: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, *63*, 456–468.

Burns, E. M. & Campbell S. L. (1994) Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: examples of categorical perception? *Journal of the Acoustical Society of America, 96*, 2704–2719.

Clough, J. & Douthett, J. (1991) Maximally Even Sets. *Journal of Music Theory, 35*, 93–173.

Cuddy, L. L., Cohen, A., & Miller J. (1979) Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology, 25*, 148–157.

Dewitt, L. A. & Crowder, R. G. (1986) Recognition of novel melodies after brief delays. *Music Perception, 3*, 259–274.

Dowling, W. J. (1991) Pitch structure. In P. Howell, R. West, & I. Cross (Eds.), *Representing Musical Structure* (33–57). London: Academic Press.

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

Dowling, W. J., Kwak, S.-Y., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics, 57,* 197–210.

Halpern, A. R. (1984) Perception of structure in novel music. *Memory and Cognition, 12*, 163–170

Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1998) Perception of mode, rhythm, and contour in unfamiliar melodies: Effects of age and experience. *Music Perception, 15,* 335–355.

Killam, R. N., Lorton, P. V., & Schubert, E. D. (1975) Interval recognition: Identification of harmonic and melodic intervals. *Journal of Music Theory*, *19*, 212–233.

Jordan, D. S., & Shepard, R. N. (1987) Tonal schemas: Evidence obtained by probing distorted musical scales. *Perception and Psychophysics, 41,* 489–504.

Leaver, A. M. & Halpern, A. R. (2003) Effects of training and melodic features on mode perception. *Music Perception, 22*, 117–143.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957) The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 61*, 358–368.

Lindblöm, B. (1996) Role of articulation in speech perception: Clues from production. *Journal of the Acoustical Society of America, 99*, 1683–1692.

McFadden, D. & Callaway, N. L. (1999) Better discrimination of small changes in commonly encountered than in less commonly encountered auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 543–560.

Macmillan, N. A. (1987) Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In S. Harnad (Ed.), *Categorical Perception* (53–88). Cambridge, Cambridge University Press.

Macmillan, N. A. (2002) Signal detection theory. In H. Paschler & J. Wixted (Eds.), *Stevens' Handbook of Experimental Psychology (3rd ed.), vol. 4: Methodology in Experimental Psychology* (43–90). Hoboken, NJ, John Wiley & Sons.

Plomp R., Wagenaar, W. A., & Mimpen, A. M. (1973) Musical interval recognition with simultaneous tones. *Acoustica, 29*, 101–109.

Rosen, S. & Howell P. (1987) Auditory, articulatory, and learning explanations of categorical perception in speech. In S. Harnad (Ed.), *Categorical Perception* (113–160). Cambridge, Cambridge University Press.

Siegel, J. & Siegel, W. (1977) Categorical Perception of tonal intervals: Musicians can't tell sharp from flat. *Perception and Psychophysics, 21*, 399–407.