Proceedings of the 2007 IEEE Symposium on Computational
Intelligence in Image and Signal Processing (CIISP 2007)

Developing Metrics for Evolving a Musically Satisfying Result from a Population of Computer Generated Music Compositions

Kris Bryden
Department of Music
Iowa State University
Ames, Iowa 50011
kabryden@iastate.edu

Abstract—**Creating a computational metric of the quality and musicality of a computer generated music composition is a challenging problem. This paper describes a possible way to develop numerically based metrics for standardizing the fitness evaluation of a population of computer generated music compositions. These metrics are based on assertions by music theorists about the dynamic forces within music that provides the music with a sense of motion and phrasing. The actions of several musical parameters are ranked according to the sense of intensity they produce. This process is applied to an acoustic music composition, and the values are then plotted on a graph. The composite curve from these values give the evaluator a sense of the patterns of lower and higher degrees of intensity, thus providing a sense of how well the music provides a sense of motion and phrasing.**

I. Introduction

Creating a computational metric to evaluate the quality and musicality of a population of music compositions generated by a computer algorithm is a challenging problem. Without a computational metric of musicality, a human listening to each music composition must perform each fitness evaluation. This is demonstrated in a recent work by the author in which a human-in-the-loop evolutionary algorithm evolves a population of musical compositions. In this algorithm each composition is created from a numerical data input is interpreted using a Circular L-system interpreter [1]. Currently the fitness function for the evolutionary process for this system is the intuition of the human evaluator. Relying on intuition is problematic because the evaluation is potentially inconsistent, and it cannot be mechanized in any way.

This paper describes a possible way to develop numerically based metrics for standardizing the fitness evaluation of a population of music compositions created by a computer algorithm. These metrics are based on the assertions of several highly regarded music theorists about the dynamic forces within music that provide a sense of motion that is shaped into both small and large gestures or phrases. Aside from some of the more experimental music in recent decades, listeners tend to expect music to provide a sense of phrasing, that is, a sense of growth and decline that can be felt as building tension and subsequent release of that tension. This emotional response is the result of a process that often involves a sense of changing degree of intensity over time. After establishing how music can provide a sense of phrasing, this discussion moves to how the actions of several musical parameters can be ranked according to greater or lesser degrees of intensity that these actions produce. After determining the numeric values for several musical parameters of a piece, these values can then be plotted onto a graph. After plotting a composite from the average of the individual curves, a fitness function of an idealized composite curve can be used to evaluate this particular composite curve. Currently, this work has progressed to the point of plotting lines of differing degrees of intensity for individual parameters and a composite curve. Section 4, "Application to an Acoustic Piece of Music" describes the process of applying the metrics currently developed to the "Invention" movement from George Perle's Wind Quintet No. 4 (1984).

II. Previous Efforts

In the past twenty or so years, composers have developed computer programs that automatically produce music compositions that fit within a certain style. An early example of this tendency is the use by [2] of "predicate calculus in 1987 to develop more than 350 rules of voice-leading for creating chorales in the style of J. S. Bach" [3]. More recently, [4] used neural nets to incorporate Renaissance and Baroque style characteristics. According to [3], the neural network learns from examples of music "through a process called backpropagation." David Cope's program "Experiments in Musical Intelligence" has introduced "association nets . . . for inductive learning. "Association nets" differ from neural nets in their use of "unlimited number of interconnected nodes". In addition, although they can "chain backward as well as forward, they do not backpropagate; thus, they do not need the same type of training as neural nets. Furthermore, "association nets do not have hidden units, as do neural nets, The nodes in association nets can be accessed at an time, revealing all of their explicit and implicit associations" [3].

Others have used evolutionary processes to improve computer-generated music. An early example, "GenJam," is a genetic algorithm-based model of a novice jazz musician learning to improvise. This program uses a neural network to augment its human fitness function [5]. Bill Manaris and several others have done extensive research in this area. Recently they have developed a set of metrics using Zipf's Law to "classify music according to pleasantness as reported by human subjects" [6]. This work is further described in [7].

III. Musical Dynamics for Providing a Sense of Phrasing

Rather than judging music by its "pleasantness" or how well it fits a given repertoire, this paper proposes measuring the sense of motion to and from a sense of stability in computer generated music compositions. Much of the background for the ensuing examination of the musical dynamics that lead to a sense of motion is based on [8]. A sense of musical motion "largely, although not exclusively, depends on *change* within one or more element-successions". "Element-successions" is defined as a succession or series of one or more musical elements such as harmony, texture, or timbre. Additionally, according to [8], the usual state of motion in music "is one of directed activity—courses of change—in lines of *growth or decline* at various levels". Berry explains that growth entails "increasing intensity" or "*progression,*" whereas decline involves "subsiding intensity" or "*recession.*" Moreover, [8] considers the possibility of stasis, a chain of events that do *not* change in intensity. These modes of musical motion depend on varying degrees of intensity created by the actions of a number of musical factors under several main categories called "elements." These elements include *tonality, melody, timbre, texture, and rhythm* [8]. These musical elements act individually and in combination to create lines of increasing and decreasing intensity.

The goal of this directed motion, according to a number of other theorists is that of movement away from and toward a stable base. For instance, in a discussion about structural downbeats within a tonal context, [9] asserts "passage *from* a stable base, or from instability *toward* a base of stability/repose, is the essence of musical motion." Ref. [10] as well incorporates the notion of goal-oriented musical motion toward stability in explaining that the human mind naturally searches for "stable shapes" or the completion of patterns. For example, after hearing a melodic leap, [10] asserts the listener "wants" to hear the gap filled in. Additionally, [10] contends that stability can only occur in a context of change, explaining that repetitious events or patterns with only uniform changes "establish no points of relative stability and closure." Often pitch elements such as harmony and melody are considered for creating a sense of musical motion toward and away from a sense of stability. However, other parameters can create this sense as well. Indeed, [8] observes that along with other musical processes, the use of rhythm plays a key role in the expression of stability and flux: "Rhythm too undergoes changes with functional consequences for music's intensity scale, playing an essential and telling role in the delineation of processes of growth and decline, climax and subsidence, stability and flux."

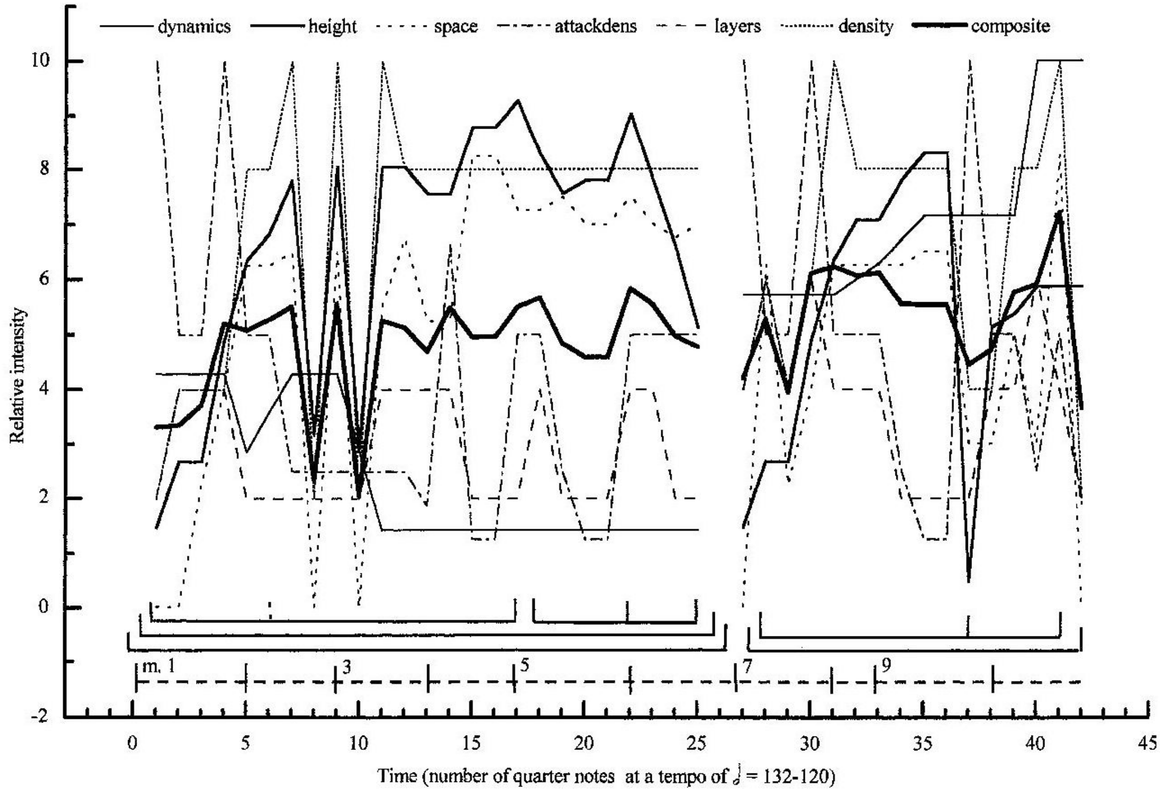
IV. Shaping Increasing and Decreasing Lines of Intensity in Several Musical Parameters

Motion in music is a felt response in the listener involving lines of increasing and decreasing intensity. When intensity increases, one senses that the music is progressive; that is, it seems to grow in emotional fervor. When intensity decreases, one senses decline or recession; in other words, the music seems to become more stable or nearer resolution. These responses are brought about by the use of certain compositional techniques within various musical factors. Outlining the tendencies of the actions of various musical parameters toward increasing or decreasing intensity is a first step toward establishing numeric values to plot on a graph. Thus, this is also a first step toward developing a fitness function based on the shape of a composite curve of the intensity curves for the individuals from a population of computer generated music compositions. Table 1 summarizes the intensifying effects of the musical parameters involved in this study, including timbre, rhythm, and texture. Under "timbre" Table 1 shows that intensity tends to increase with a greater degree of loudness and a higher register. Although these two assumptions are intuitive, they do have a physical basis. In [11] it is pointed out that greater degrees of loudness involve increasing sound pressure level. Concerning register, [10] asserts, "because they require special effort and control, we are particularly sensitive to the covert tension of high, cantabile tones." Indeed, studies have shown that listeners

TABLE 1
TENDENCIES TOWARD GREATER AND LESSER DEGREES OF INTENSITY

Timbre	Greater intensity	Lesser intensity
Dynamics	Louder	Softer
Register	Higher	Lower
Texture	More complex	Less complex
Density-number	More components sounding concurrently	Less components sounding concurrently
Texture-space	Greater musical space between lowest and highest pitches	Less musical space
Rhythmic interaction	Diverse → Similar →	Same
Rhythm	Faster rate	Slower rate
Attack frequency (density)	More frequent	Less frequent

Fig. 1. Intensity curves for mm. 1-10 of George Perle's "Invention."



associate higher register with increased physical effort. Ref. [12] cites an experimental study in [13] that backs up this assertion by showing that the subjects associated a "greater degree of muscular or sensory tension" with higher pitches [12]. Performers also experience this phenomenon. For example, there is a large degree of muscular tension involved in producing the highly focused flute embouchure to produce high notes.

Under "Texture" in Table 1, "Density-number" refers to the number of simultaneously sounding instruments or voices. As shown in Table 1, there tends to be increased intensity as the number of components rises because of the resulting thicker texture. Table 1 also summarizes tendencies for increasing or decreasing intensity with texture space (interval between the highest and lowest pitches sounding at the same time). In support of this notion, Ref. [14] cites studies in [15] and [16] that show "well-spaced frequencies" sound louder than frequencies that are next to each other. Table 1 shows that musical events that come at a faster rate tends toward greater intensity than those that come at a slower rate. Quicker rates of change tend to increase intensity by bringing about a sense of agitation and movement. Attack frequency in Table 1 refers to the number of notes being articulated within a given amount of time. Attacks that are more frequent generally create greater intensity than less frequent attacks because musical events arrive at a faster rate, thus providing a sense of greater growth and excitement. Ref. [9] provides an example of this phenomenon when he observe that Brahms's music often increases the "density of events" (by increasing the number of notes per beat) to create "a sense of accelerando" or speeding up without changing the duration of the beat. Decreasing the density of events has the same result as a *ritardando* or slowing down.

V. Proposed Numerical Metrics

Assigning numeric values to an intensity scale for each of the various musical parameters discussed above enables easy comparison between them and provides a means to measure the composite effect of all of the musical parameters combined. This information has the potential to facilitate the development of a standardized fitness function to evolve a population of music compositions created by a computer algorithm. Placing the results in a graph can enable easy comparison between the compositions. Graphs highlight processes within individual musical factors and facilitate comparison between them. For example, it is quicker and easier to interpret a rising line as a crescendo than to have to interpret a series of symbols (***p, mp, mf, f*** etc.) to achieve the same cognitive result. Additionally, graphs show the wave-like structures created by differing degrees of intensity, vividly showing patterns of musical growth, decline, and stasis. A

TABLE 2
NUMERICAL VALUES FOR THE INTERSITY CURVE "ATTACKDENS"

𝅘𝅥𝅯 96	♪ 48	♩ 24	𝅗𝅥. 9
3 : 2 ♪ 64	3 : 2 ♩ 32	𝅗𝅥 12	

fitness function could then be developed to evaluate the shape of these curves.

VI. Application to an Acoustic Piece of Music

To demonstrate the process, I have created a graph using the above metrics using data points from mm. 1-10 of George Perle's "Invention" from his *Wind Quintet No. 4*. Fig. 1 contains the graph for mm. 1-10, and Fig. 2 and 3 contains the score. From the score of the entire movement, I extracted quantitative data to construct intensity curves for dynamics, pitch height, (textural) space, attackdens (frequency of attack), density, and layers (rhythmic interaction between lines). To construct this graph, I normalized each component so that all the intensity data points range from zero to ten in order to facilitate the examination of fundamental trends and the comparison between factors. I then assigned one data point for each quarter note beat of mm. 1-10 of Perle's Invention. Additionally, I constructed a composite intensity curve to show the composite effect of the individual curves. To construct the composite curve, I added the data from each individual curve for each quarter-note duration and divided the result by five, the number of individual curves. To normalize the values for the individual intensity curves, I divided each data point by the difference between the highest and lowest data points within the entire movement for each intensity curve and then multiplied by ten. For example, the data for dynamics represents *pianissimo* (very soft) through a triple *forte* (extremely loud) (***pp*** = 1, ***p*** = 2, ***mp*** = 3, ***mf*** = 4, ***f*** = 5, ***ff*** = 6, and ***fff*** = 7). To obtain each normalized data point $dynamics_n$ for the dynamics intensity curve, I used the following equation for each "real" data point $dynamics_r$:

$$dynamics_n = (dynamics_r / 7) * 10. \quad (1)$$

Fig. 2. Score for mm. 1-3 of George Perle's "Invention."

To determine the normalized data points for the height intensity curve, I first determined the lowest and highest points of the upper voice in the movement—an F3 in m. 76 (7 semitones below middle C) and a B6 in m. 57 (35 semitones above middle C). Then I added 7 to each data point to start all the numbers at zero and then divided each by 42, the highest pitch plus 7. Then I multiplied each point by 10. The following equation summarizes this operation ($height_n$ represents each normalized data point for the height intensity curve, and $height_r$ represents each "real" data point):

$$height_n = ((height_r + 7) / 42) * 10. \quad (2)$$

To calculate each normalized data point for the space intensity curve, I first determined that the highest number of semitones between the highest and lowest pitches in a given moment is 40 in m. 33. At this moment, the flute plays an E-flat6 and the horn plays a B2. Then to obtain each normalized data point $space_n$ for the space intensity curve, I used the following equation for each "real" data point $space_r$:

Fig. 3. Score for mm. 4-10 of George Perle's "Invention."

$$space_n = (space_r \ / 40) * 10. \quad (3)$$

For this movement, each data point for "attackdens" represents a quarter note at the tempo =132. I used the numerical values for the note values shown in Table 2.

To obtain each normalized data point $attackdens_n$ for the attackdens intensity curve, I used the following equation for each "real" data point $attackdens_r$:

$$attackdens_n = (atackdens_r \ / 96) * 10. \quad (4)$$

The highest number of rhythmically differentiated melodic lines in this movement is four. To obtain each normalized data point $layers_n$ for the layers intensity curve, I used the following equation for each "real" data point $layers_r$:

$$layers_n = (layers_r / 4)* 10. \quad (5)$$

The highest number of voices playing at one time is five. To obtain each normalized data point $density_n$ for the density intensity curve, I used the following equation for each "real" data point $density_r$:

$$density_n = (density_r / 5) * 10. \quad (6)$$

The composite intensity curve in the opening seven measures before the rest shown by a space in the graph, features a series of deep fluctuations in mm. 2-3 and more broad, shallower oscillations in mm. 3-6. In the second large musical gesture, after the rest, the composite curve shows an introductory brief shallow curve followed by two well-developed curves.

In the opening small gesture in mm. 1-2 (delineated by a dotted line in the highest set of brackets beneath the intensity curves), the composite intensity begins with a large rise from 3.3 to 5.5. Large ascents in the height, space, and density intensity curves account for this rise. The height intensity curve rises significantly from 1.5 to 7.8, reflecting a greater than two octave ascent from B3 in the opening clarinet motive to a C-sharp 6 at the end of the rising flute gesture in m. 2. The space intensity curve begins at intensity level 0, representing the opening lone clarinet motive. This curve quickly rises to 6.3 by the downbeat of m. 2 due to the increasing space between the bassoon and the upper voice (the oboe followed by the flute). The intensity curve for density depicts a rise in intensity levels from 2 to 10, reflecting the consecutive entrances of all instruments in the quintet.

The other intensity curves—dynamics, layers, and attackdens—are more independent in mm. 1-2. The dynamics intensity curve remains at 4.3 through m. 1, representing a mezzo piano, and then drops to a 2.9 on the downbeat of m. 2, representing a piano in all sounding voices. This curve then returns to 4.3 on the third beat, representing a small crescendo in all the parts. This crescendo contributes to the general rise in intensity in m. 2. The intensity curve for layers rises from 2 to 4 in m. 1 because the oboe and bassoon enter with contrasting rhythms. In m. 2 this curve moves back down to 2 because all sounding voices (oboe and bassoon plus the entering flute and clarinet) articulate the same rhythm, a series of staccato eighth notes. The attackdens intensity curve swings between 10 and 5 in this initial gesture, reflecting a mixture of sixteenth and eighth notes.

The following small gesture in mm. 2-4 begins with widely fluctuating intensity levels that subsequently stabilize. In mm. 2-3 the individual intensity curves for height, space, and density feature deep fluctuations. "Height" quickly fluctuates between low points 3.2 and 2.9 and high points at 8. The lower points of this wide fluctuating curve represent an F-sharp4 and F4 in the lone horn and the high points stand for a D6 in the flute. The space and density intensity curves also include wide fluctuations because only one instrument plays during the low moments and the full quintet enters with the high flute. On the third beat of m. 3, the composite intensity curve begins a narrower range of oscillations that are more spread out. Less dramatic movement in the height intensity curve and the near constant level of the density intensity curve explain this difference. The highest sounding voice—the flute—stays within an octave, rising from C6 to F6 in m. 4, shown by the height intensity level ascending from 7.6 to 9.3. On the fourth beat of m. 3, the intensity curve for density drops from 10 to a sustained level 8 because the horn withdraws and the rest of the quintet continues to play.

In mm. 5-6, the composite intensity curve outlines two smaller gestures (one in m. 5 and one in m. 6) that are also delineated by the top brackets in Fig. 1. The individual intensity curves for height and layers have the same pattern as the composite intensity curve in both these gestures. For the small gesture in m. 5, "attackdens" also has the same shape, and in m. 6, the space intensity curve possesses the same shape as the composite intensity curve. The intensity curve for dynamics and density do not change in mm. 5-6, but the instrumentation does change slightly. The bassoon rests and the horn comes in.

The composite intensity curve shows that the second phrasal gesture in mm. 7-10 has a more defined shape than the first in mm. 1-6. After the opening quick ascent and descent, it outlines two smaller gestures (one in mm. 7-9 and one in mm. 9-10) delineated by the upper divided bracket under the intensity curves in Fig. 1. For the first of these smaller gestures, the composite intensity curve shows a consolidated ascent to 6.2 on the downbeat of m. 8 and then slowly descends to 4.4 for the fifth beat of m. 9. Longer note values and a simpler texture provide the ending of this gesture with a partial sense of stability.

The closing gesture, beginning on the fourth beat of m. 9, is quick, straightforward, and emphatic. The composite intensity curve begins this gesture at 4.4 and rises to 7.2 on the fourth beat of m. 10, the highest level of intensity thus far in the movement. This curve then ends with a quick descent to 3.6 on the fifth beat of m. 10. The intensity curves for layers and density have a similar contour as the composite curve, and both the space and attackdens intensity curves ascend and descend twice during this final gesture. The layers intensity curve rises from 2 to 6 on the third beat of m. 10, depicting an increase in rhythmic complexity. Intensity levels decrease through beats 4 and 5 as rhythmic diversity decreases. (All sounding voices except the flute play two eighth notes on beat 4, and only the flute plays on beat 5.) The intensity curve for density rises from 4 to 10 from the fifth beat of m. 9 to the fourth beat of m. 10 as more voices enter in. This curve drops significantly on the fifth beat of m. 10 where the flute is the only remaining voice. Unlike the other intensity curves, dynamics and height ascend throughout this final gesture. The dynamics intensity curve remains at 7.1 until the third beat of m. 10 where it rises to 10, representing a triple *fortissimo* (extremely loud) in the flute and *sforzando fortissimos* (suddenly loud) in the clarinet and bassoon. The height intensity curve shows a continuous rise from 0.5 (representing the G3 beginning this final gesture in the clarinet) to a 5.9 (standing for the F4 in the flute).

VII. CONCLUSIONS AND FUTURE WORK

The next step in this work is to develop criteria for well-formed composite curves that represent a musical result. Once this is complete, the system is ready to be used to supplement the intuition of a human evaluator and possibly to facilitate automating the entire evolutionary algorithm. Some factors that could be considered are the time span of waves of the composite curve and their depth, proportions, and shape. Then this methodology could be applied to a population of computer generated music compositions. Applying this would be a three step process: first to establish numeric data about the various musical parameters described above, second to plot the results on a graph, and third to assign a fitness value to the composite curve. After accomplishing these tasks, further work could be done to automate the process. This work would include writing software to ascertain the numerical data for each computer generated music composition using the MIDI information from a computer algorithm. Developing software to plot the results and evaluate the resulting composite curves would automate the process further.

The musical parameters used in this work can potentially be expanded. Other textural parameters could include "density-compression," where intensity is likely to increase as more components are packed into a "given total space" [8]. "Distribution," is a subcategory of "density-compression," and entails the intervallic distribution of the components within the given total space. Ref. [8] observes that a more dissonant PC or IC content has an intensifying influence on the texture. Other aspects of rhythm could be plotted also. One aspect could include tempo, which is the rate of change for musical events. Another rhythmic aspect that could be developed is how complex the metrical relationships are. For example, polymeter, where simultaneously sounding parts have different meters would have a higher intensity rating than if all the parts were in the same meter. Additionally, asymmetrical meters would create a greater sense of intensity than symmetrical meters.

REFERENCES

[1] K. Bryden, "Using a Human-in-the-Loop Evolutionary Algorithm to Create Data-Driven Music." In *Proceedings of the 2006 IEEE World Congress on Computational Intelligence,* Vancouver, CA, 2006.

[2] K. Ebcioglu, *CHORAL Project: An Expert System for Harmonizing Four-Part Chorales.* Yorktown Heights, NY: IBM Thomas J. Watson Research Center, 1987.

[3] D. Cope, *Virtual Music: Computer Synthesis of Musical Style,* Cambridge, MA: MIT Press, 2001.

[4] D. Hörnel and W. Menzel, "Learning Musical Structure and Style with Neural Networks." *Computer Music Journal* 22/4: 44-62, 1998.

[5] J.A. Biles, P.G. Anderson, and L.W. Loggi, "Neural Network Fitness Functions for a Musical IGA," in *Proceedings of the IIA '96/SOCO '96, International Symposium on Intelligent Industrial Animation and Soft Computing, B39-B44, 1996.*

[6] B. Manaris, P. Machado, C. McCauley, N. Romero, and D. Krehbiel, "Adaptive Critics for Evolutionary Artists," EvoMusart 2005, *Lecture Notes in computer Science Applications of Evolutionary Computing, LNCS 3449,* Springer-Verlag, 437-447, 2005.

[7] B. Manaris, et al., "Zipf's Law: Music Classification and Aesthetics," *Computer Music Journal*, 29/1, 2005.

[8] Berry, Wallace, *Structural Functions in Music,* New York: Dover, Reprint of 1976 edition with corrections by the author, Englewood Cliffs, NJ: Prentice-Hall, 1987.

[9] D. Epstein, *Shaping Time: Music, the Brain, and Performance,* New York, NY: Schirmer Books, 1995.

[10] L.B. Meyer, *Explaining Music,* Chicago, IL: The University of Chicago Press, 1973.

[11] R.A. Rasch and R. Plomp, "The Perception of Musical Tones," In *The Psychology of Music,* ed. Diana Deutsch. Orlando, FL: Academic Press, 1-24, 1982.

[12] R. Francès, translated by W. J. Dowling, *The Perception of Music*, Hillsdale, NJ: Lawrence Erlbaum Assoc., 1988.

[13] G. Révész, Gibt es ein Hörraum [Is there spatial hearing?], *Acta Psychologica* 3, 137-192, 1939.

[14] R.E. Radocy and J.D. Boyle, *The Psychological Foundations of Musical Behavior,* 3rd ed. Springfield, IL: Charles C. Thomas, 1997.

[15] S.S. Stevens, "The Relation of Pitch to Intensity," *Journal of the Acoustical Society of America* 6, 150-154, 1956.

[16] E.G. Zwickler, G. Flottorp, and S.S. Stevens, "Critical Band Width in Loudness Summation," *Journal of the Acoustical Society of America* 29, 548-557, 1957.