De-mystifying Music and Its Performance

Elaine Chew
University of Southern California
Andrew and Erna Viterbi School of Engineering
Flora L. Thornton School of Music

The mathematical nature of music and its imminently quantifiable attributes make it an ideal medium through which to study human creativity and cognition. The architecture of musical structures reveal the principles of invention and design; the dynamics of musical ensemble offer models of human collaboration; and, the demands of musical interaction challenge existing and require new computing paradigms.

Engineering methodology forms an integral part of systematic study, computational modeling, and scientific explanations of human abilities in music perception and cognition, and in music making. In turn, better understanding of the in-action thinking and problem solving innate in music making and cognition can provide valuable insights into the mechanics of engineering discovery and design. Commercial interest in Internet radio, music recommendation and discovery, and video games further fuels the proliferation of engineering-music research.

This paper presents some projects originating from the Music Computation and Cognition Laboratory at the University of Southern California – representing research in computational music cognition, expression synthesis, ensemble coordination, and musical improvisation – to give a gist of the richness and breadth of research possible at the intersection of engineering and music performance.

1. MuSA.RT – Analyzing and Visualizing Tonal Structures in Real Time

Tonal music, which refers to almost all of the music that we hear, consists of tones (or pitches) organized in time that generate the perception of different levels of stability. The most stable tone, the key, serves as the tonal center of the musical segment. The computational modeling of key finding dates back to the early days of artificial intelligence (Longuet-Higgins & Steedman 1971). A popular key finding method was devised by Krumhansl and Schmuckler (Krumhansl 1990), and based on the computing of correlation coefficients between the duration profile (vector) of a query stream and experimentally derived probe tone profiles.

More recently, Chew (2000) proposed the spiral array model for tonality, which consists of a series of nested helices representing pitch classes, and major and minor chords and keys. The representations are generated by successive aggregation of their component parts. Correspondences exist between the spiral array and prior models (Chew 2008). The spiral array lends itself readily to the design of efficient algorithms for automated tonal analysis, and to the scientific visualization of these algorithms and of musical structure.

Any stream of notes can generate a center of effect (*c.e.*), a center of gravity of the notes, in the spiral array space. The Center of Effect Generator (CEG) key finding algorithm based on the spiral array searches for the key representation nearest to the c.e. of the query stream to determine key. The interior point approach of the CEG algorithm speeds the recognition of key and provides a framework on which to design further algorithms for automated music analysis.

A natural extension of the key finding problem is the search for key (or contextual) boundaries. Two algorithms have been proposed to find key boundaries using the spiral array: one that minimizes the distance between each segment's CE and its closest key (Chew 2002); and, another that finds statistically significant maxima in the distance between the CE's of consecutive music segments, without regard to key (Chew 2005).

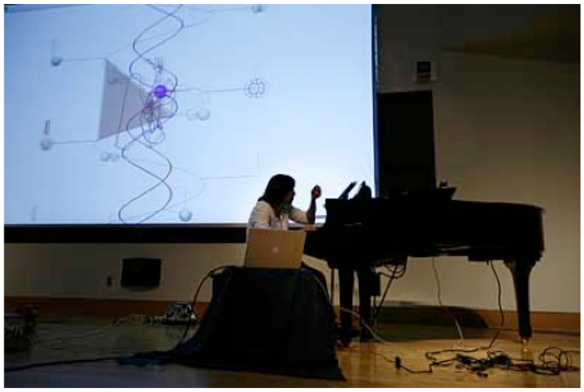


Figure 1. MuSA.RT in concert at the 2008 ISMIR conference at Drexel University
© The Philadelphia Enquirer

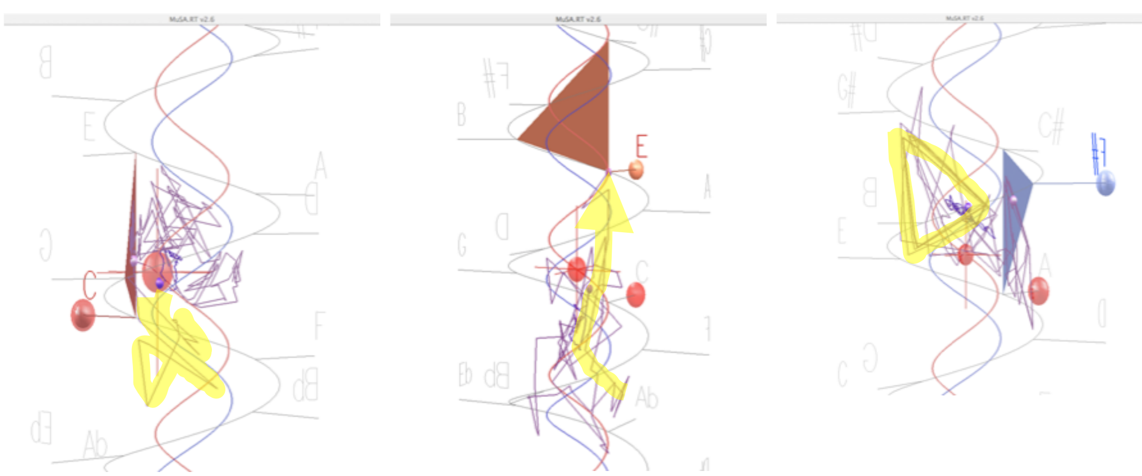
Essential to automated music transcription is the inverse problem of pitch spelling – turning note numbers or frequency values into letter names for human-readable music manuscripts. Several variants of a pitch-spelling algorithm using the spiral array have been proposed: a cumulative window (Chew & Chen 2003a), a sliding window (Chew & Chen 2003b), and a multi-window bootstrapping algorithm (Chew & Chen 2005).

Converting theoretically efficient algorithms into robust working systems that stand up to the rigors of performance presents a different set of challenges. Using the Software Architecture for Immersipresence framework (François to appear), many of the above algorithms were incorporated into MuSA.RT – Music on the Spiral Array . Real-Time – an interactive tonal analysis and visualization software (Chew & François 2003).

MuSA.RT has been used in analysis and visualization of music by Pachelbel, Bach, and Barber (Chew & François 2005), in juxtaposition to Sapp's keyspaces (2005) and Toiviainen's self-organizing maps (2005). MuSA.RT has also been demonstrated in concerts internationally and at the AAAS Ig Nobel session in 2008. Figure 1 shows MuSA.RT in concert at the 2008 ISMIR conference at Drexel University in Philadelphia.

In MuSA.RT v.2.7, the pitch class helix and pitch names are shown in silver, the major/minor triad helices are hidden, and the major/minor key helices are shown in red/blue, respectively, and appear as a double helix in the structure's core. Silver spheres appear on a note name when it is sounded. A short-term CE tracks the local context, and the triad closest to it lights up as a pink/blue triangle; a long-term CE tracks the larger scale context, and a sphere appears on the closest key, the size of which is inversely proportional to the CE-key distance. Violet and indigo trails trace the history of the short-term and long-term CE's, respectively.

Huron's analysis (2004) of musical humor devices employed by P.D.Q. Bach



(a) jazz ending in Prelude I (b) unusual tonal shift in Fugue II (c) excessive repetition in Prelude X

Figure 2. Humor devices (highlighted in yellow) in P.D.Q. Bach's *Short-Tempered Clavier* visualized in MuSA.RT (adapted from figures in Chew & François 2009)

Figure 3. ESP at the USC Festival 125 Pavilion. Photo by EC.

(a.k.a. Peter Schickele) revealed that many involved violation of expectations. We used MuSA.RT to see incongruous styles, improbable harmonies, surprising tonal shifts (all of which appear as far-flung trajectories in the spiral array space), and excessive repetition in P.D.Q. Bach's *Short-Tempered Clavier* (Chew & François 2009). Figure 2 shows a few examples of visualizations of such humor devices.

By judicious use of the sustain pedal, and by accenting different notes through duration or stress, a performer can guide the perception of tonal structures. The latest versions of MuSA.RT take into account pedal effects in the computing of tonal structures (Chew & François 2008). Because the CE trails react directly to the timings of note soundings, no two human performances of the same piece will result in the same trajectories.

Ongoing work include extending the spiral array concepts to a higher dimension to uniquely represent tetrachords (Alexander et al. in preparation). The resulting pentahelix has direct correspondence to the orbifold model (Tymoczko 2006, Callendar, Quinn & Tymoczko 2008).

2. ESP – Experiencing Music Performance through Driving

Not everyone can play an instrument well enough to be able to execute expressive decisions and interpretations at will, but almost anyone can drive a car, at least one in a simulation. ESP, the Expression Synthesis Project, takes the musical analogy of locomotion literally to create a driving interface for expressive performance so that novices can experience the kind of embodied cognition characteristic of expert musical performance.

In ESP (Chew et al. 2005), the user can increase or decrease the speed of the car (music playback) using the accelerator and the brake pedal. The road's centerline segments approach at one per beat, giving a sense of the tempo (beat rate and car velocity), which is shown in the speedometer in beats per minute. Suggestions to slow

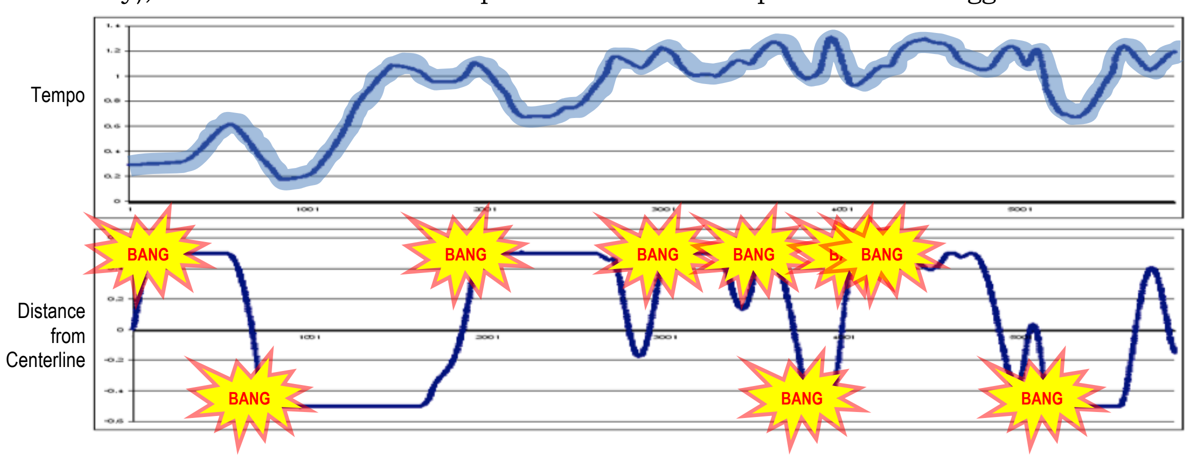


Figure 4. Virtual radius mapping ensures smooth tempo despite erratic driving (adapted from figure in Liu, Chew & François 2006)

down or speed up are embedded in the road bends and straight sections, respectively. The road map thus corresponds to an interpretation, and often reveals the underlying structure of the piece.

Despite the expressive suggestions, the user is free to choose her/his desired tempo trajectory. Also, more than one road map (interpretation) can correspond to the same piece of music (Chew et al. 2006). As part of the system design (using SAI), a virtual radius mapping strategy ensures tempo smoothness, a hallmark of expert performance, even in extreme cases of erratic driving behavior (Liu et al. 2006), as shown in Figure 4.

3. DIP – Charting the Dynamics of Ensemble Coordination

Remote collaboration is an integral part of our increasingly global and distributed workplace and society. Musical ensemble performance offers a unique framework through which to examine the dynamics of close collaboration and the challenges of human interaction in the face of network delay.

In a series of Distributed Immersive Performance (DIP) experiments, we recorded the Tosheff Piano Duo performing Poulenc's *Piano Sonata for Four Hands* under auditory delay ranging from 0ms to 150ms. Self reports showed that, beyond 50ms, the players struggled to keep time, and interpretation was comprised (Chew et al. 2004), in a setting where they heard themselves immediately, and their partners with delay. By delaying each pianist's playing to their own ears so that it aligned with the incoming signal of their partner's playing (see Figure 5), we were able to create a more satisfying experience for the musicians that allowed them to adapt to the conditions of delay, and to increase the threshold to 65ms (Chew et al. 2005), even for the fast and rhythmically demanding *Final*. Segmental tempo analysis (Chew et al. 2005) showed a marked increase in the range of segmental tempo strategies between 50ms and 75ms, and a decline at 100ms and 150ms.

Few other experiments exist that treat network delay as a constraint, and not

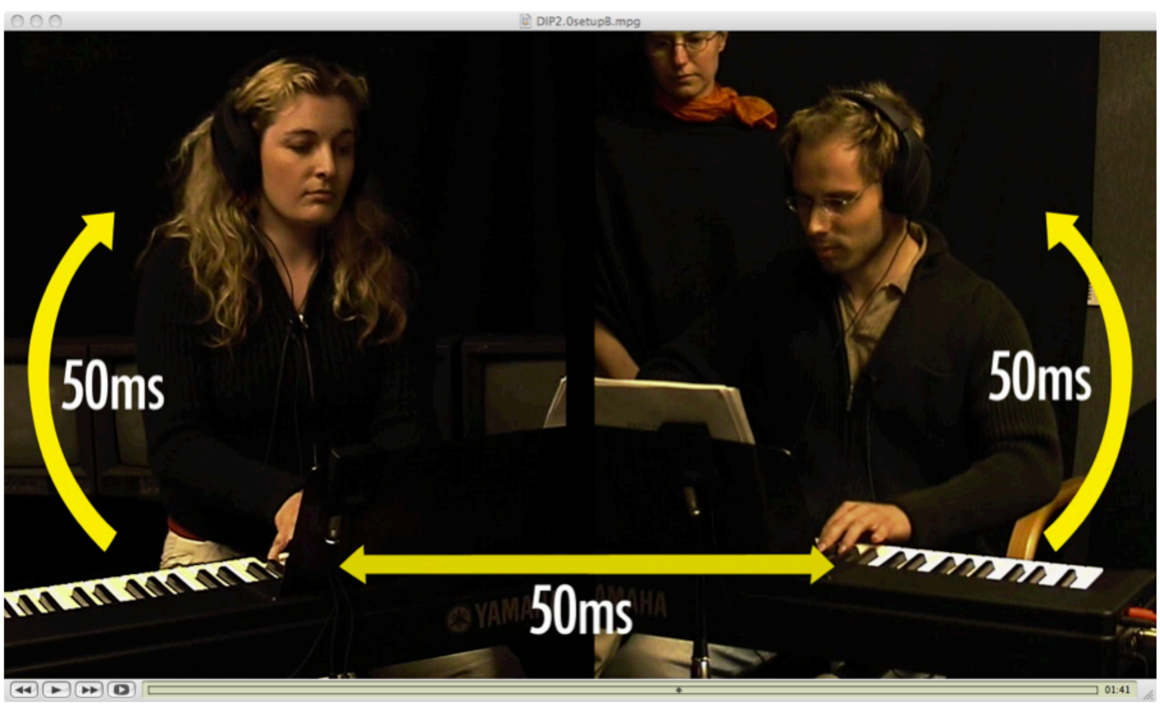


Figure 5. Solution to network delay: delaying each player's sound to their own ears to align with incoming audio of their partner's playing

simply a feature for free improvisation. A clapping experiment at Stanford showed that pairs of musicians slowed down over time with increasing auditory delay, and sped up modestly for delays below 11.5ms (Chafe & Gurevich 2004). Similar experiments at the University of Rochester (Bartlette et al. 2006) found that latencies above 100ms profoundly impacted musicians' abilities to play as a duo. More recent tools and techniques for music alignment and performance analysis are allowing us to conduct further examinations of the DIP files towards the creating of detailed maps of ensemble dynamics (Wolf & Chew 2010).

4. MIMI – On-the-fly Architecting of a Musical Improvisation

MIMI, Multimodal Interaction for Musical Improvisation, was created as a standalone performer-centric tool for human-machine improvisation (François, Chew & Thurmond 2007). Figure 6 shows MIMI at her concert debut earlier this year.

MIMI takes user input, creates a factor oracle, and traverses it stochastically to generate recombinations of the original input (Assayag & Dubnov 2004). In existing improvisation systems (Assayag et al. 2006, Pachet 2003), performers react to machine output without prior warning. MIMI allows for more natural human interaction by providing visual cues to the origins of the music generated, and by giving a ten-second heads up and review of the music material.

The MIMI system allows the performer to decide when the machine learns and when the learning stops, and to determine the recombination rate (the degree of fragmenting of the original material), the loudness of playback, and when to clear the memory (François, Schankler & Chew 2010). By tracking the decisions over the unfolding of a performance, we build a record of the architecting of an improvisation. Ongoing work includes the documenting of such performance decisions so as to better understand musical design.

5. Open Courseware

Further reviews of music-engineering topics exist as open courseware at *www-scf.usc.edu/~ise575* (Chew 2006). Each website contains a reading list, presentations and students' reviews of the papers, and links to final projects. For the 2010 topic on Musical Prosody and Interpretation, student project highlights include Brian Highfill's time warping of a MIDI file of *Wouldn't It Be Nice* to align with the Beach Boys' recording of the piece, Chandrasekhar Rajagopal's comparison of guitar and piano performances of Granados' *Spanish Dance No. 5*, and Balamurali

Figure 6. MIMI's concert debut at the People Inside Electronics concert in Pasadena, California, with Isaac Schankler. Photo by EC.

Ramasamy Govindaraju's charting of the evolution of vibrato in violin performance over time.

6. Acknowledgments

This work was supported in part by the National Science Foundation (NSF) under Grant Nos. 0321377 and 0347988, and Cooperative Agreement No. EEC-9529152. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of NSF.

7. References

Longuet-Higgins, H. C., and M. Steedman (1971). "On Interpreting Bach." *Machine Intelligence* 6:221-241.

Krumhansl, C. (1990). *Cognitive Foundations of Musical Pitch.* Oxford, UK: Oxford University Press.

Chew, E. (2000). *Towards a Mathematical Model of Tonality*. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2000

Chew, E. (2008). Out of the Grid and Into the Spiral: Geometric Interpretations of and Comparisons with the Spiral-Array Model. W. B. Hewlett, E. Selfridge-Field, and E. Correia, Jr. (eds.): *Computing in Musicology* 15: 51-72.

Chew, E. (2002). The Spiral Array: An Algorithm for Determining Key Boundaries. In C. Anagnostopoulou, M. Ferrand, A. Smaill (eds.): *Music and Artificial Intelligence*, Springer LNCS/LNAI, 2445:18-31.

Chew, E. (2005). Regards on Two Regards by Messiaen: Post-tonal Music Segmentation Using Pitch Context Distances in the Spiral Array. *Journal of New Music Research*, 34(4):341-354.

Chew, E. & Y.-C. Chen (2003a). Mapping MIDI to the Spiral Array: Disambiguating Pitch Spellings. In Proceedings of the 8th INFORMS Computing Society Conference, Chandler, AZ, Jan 8-10, 259-275.

Chew, E. and Y.-C. Chen (2003b). Determining Context-Defining Windows: Pitch Spelling using the Spiral Array. In *Proceedings of the 4th International Conference for Music Information Retrieval*, Baltimore, MD, Oct 26-30, 2003.

Chew, E., and Y.-C. Chen (2005). Real-Time Pitch Spelling Using the Spiral Array. *Computer Music Journal*, 29(2):61-76.

François, A. R. J. (to appear). An Architectural Framework for the Design, Analysis and Implementation of Interactive Systems. *The Computer Journal.*

Chew, Elaine & François, Alexandre R. J. (2003). MuSA.RT - Music on the Spiral Array . Real-Time. In *Proceedings of the ACM Multimedia '03 Conference*, Berkeley, CA, Nov 2-8, 448-449.

Chew, Elaine and François, A.R.J. (2005). Interactive multi-scale visualizations of tonal evolution in MuSA.RT Opus 2. *ACM Computers in Entertainment*, 3(4):16 pages.

Sapp, C. (2005). Visual hierarchical key analysis. *ACM Computers in Entertainment*, 3(4):19 pages.

Toiviainen, P. (2005). Visualization of tonal content with self-organizing maps and self-similarity matrices. *ACM Computers in Entertainment*, 3(4):10 pages.

Huron, D. (2004). Music-engendered laughter: an analysis of humor devices in P.D.Q. Bach. In *Proceedings of the International Conference on Music Perception and Cognition*, Evanston, IL, 700-704.

Chew, E. and A. R. J. François (2009). Visible Humour - Seeing P.D.Q. Bach's Musical Humour Devices in The Short-Tempered Clavier on the Spiral Array Space. In T. Noll, T. Klouche (eds.): *Mathematics and Computation in Music - First International Conference*, Springer CCIS 37.

Chew, E. and A. R. J. François (2008). MuSA.RT and the Pedal: the Role of the Sustain Pedal in Clarifying Tonal Structure. In *Proceedings of the Tenth International Conference on Music Perception and Cognition*, Sapporo, Japan, August 25-29.
Alexander, S., E. Chew, R. Rowe, S. Rodriguez (in preparation). The Pentahelix: A Four-Dimensional Realization of the Spiral Array.
Tymoczko, D. (2006). The Geometry of Musical Chords. *Science* 313(5783):72-74.
Callendar, C., I. Quinn and D. Tymoczko (2008). Generalized Voice-Leading Spaces. *Science* 320(5874):346-348.
Chew, E., A. R. J. François, J. Liu and A. Yang (2005). ESP: A Driving Interface for Musical Expression Synthesis. In *Proceedings of the Conference on New Interfaces for Musical Expression*, Vancouver, B.C., Canada, May 26-28.
Liu, J., E. Chew, and A. R. J. François (2006). From Driving to Expressive Music Performance: Ensuring Tempo Smoothness. In *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, Hollywood, CA, June 14-16.
Chew, E., J. Liu, and A. R. J. François (2006). ESP: Roadmaps as Constructed Interpretations and Guides to Expressive Performance. In *Proceedings of the First Workshop on Audio and Music Computing for Multimedia*, Santa Barbara, CA, Oct 27, 137-145.
Chew, E., R. Zimmermann, A. A. Sawchuk, C. Kyriakakis, C. Papadopoulos, A. R. J. François, G. Kim, A. Rizzo and A. Volk (2004). Musical Interaction at a Distance: Distributed Immersive Performance. In *Proceedings of the 4th Open Workshop of MUSICNETWORK*, Barcelona, Spain, Sep 15-16.
Chew, E., R. Zimmermann, A. Sawchuk, C. Papadopoulos, C. Kyriakakis, C. Tanoue, D. Desai, M. Pawar, R. Sinha and W. Meyer (2005). A Second Report on the User Experiments in the Distributed Immersive Performance Project. In *Proceedings of the 5th Open Workshop of MUSICNETWORK*, July 4-5, 2005, Vienna, Austria.
Chew, E., A. Sawchuk, C. Tanoue and R. Zimmermann (2005). Segmental Tempo Analysis of Performances in Performer-Centered experiments in the Distributed Immersive Performance Project. In *Proceedings of International Conference on Sound and Music Computing*, Salerno, Italy, November 24-26.
Chafe, C. and M. Gurevich (2004). Network Time Delay and Ensemble Accuracy: Effects of Latency, Asymmetry. In *Proceedings of the 117th Conference of the Audio Engineering Society*, San Francisco, California.
Bartlette, C., D. Headlam, M. Bocko and G. Velikic (2006). Effect of Network Latency on Interactive Musical Performance. *Music Perception* 24(1):49-62.
Wolf, K. and E. Chew (*2010).* Evaluation of Performance-to-Score MIDI Alignment of Piano Duets. In online abstracts of the International Conference on Music Information Retrieval, August 9-13, Utrecht, the Netherlands.
François, A. R. J., E. Chew and C. D. Thurmond (2007). Visual Feedback in Performer-Machine Interaction for Musical Improvisation. In *Proceedings of International Conference on New Instruments for Musical Expression*, New York, June.
Assayag, G. and S. Dubnov (2004). Using Factor Oracles for machine Improvisation. *Soft Computing* 8(9):604-610.
Assayag, G., G. Bloch, M. Chemillier, A. Cont and S. Dubnov (2006). Omax Brothers : a Dynamic Topology of Agents for Improvization Learning. In *Proceedings of the Workshop on Audio and Music Computing for Multimedia*, Santa Barbara, CA.
Pachet, F. (2003). The continuator: Musical interaction with style. *Journal of New Music Research* 32(3):333-341.
François, A. R. J., I. Schankler and E. Chew (2010). Mimi4x: An Interactive Audio-visual Installation for High-level Structural Improvisation. In *Proceedings of the Workshop on Interactive Multimedia Installations and Digital Art*, Singapore, July 23.
Chew, E. (2006). Imparting knowledge and skills at the forefront of interdiscinterdisciplinary research - a case study on course design at the

intersection of music and engineering. In *Proceedings of the 36th Annual ASEE/IEEE Frontiers in Education Conference*, San Diego, October 28-31.