ATMM 2014
PROCEEDINGS
Audio Technologies for Music and Media
International Conference

Bilkent University
Faculty of Art, Design and Architecture
Department of Communication and Design
Bilkent, 06800 Ankara, Turkey

Phone: +90-312-290-1749

www.atmm-conference.org
# CONTENTS

**Introduction**

Creative Use of Synthestration and MIDI Virtual Instruments in a Jazz Recording and Film Scoring Context  
Daniel A. Walzer  
7

Sparse Solutions in Audio Inpaintinga  
Mehmet Erdal Özbek  
14

Precis for Celtic Music Radio (CMR) Final  
Patrick Quinn  
26

A Transmedial Approach to the Distribution of Music in Online Environments: aThe Hype Machine  
Buğu Melis Çağlayan  
30

Messiaen, Composer of ‘Awe’: The Proto-Spectralist Pursuit of Divine Fear in Apparition de l'Église éternelle  
Miriam Pilonen  
36

Establishing a Signal Processing Methodology for Multi-track Mixing Education  
C. Barkın Engin  
47

You Are a Target Market: Using News Media and Advertising to Create Generative Sound and Visual Art  
Alayna Aine Hughes  
57

A Survey Of Pitch Correction Methods  
Zeki Hayran, Ali Burak Parım, Mehmet Saygı̇n Seyfiö̈ğlu  
64

Music and Audio Production for Documentary Films “Over the example of a particular documentary: Yılan Hikayeleri (Snake Stories)”  
Cihan Işıkhan  
79
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Current Challenges Facing the Sound Designers and Music Composers for Film, and Modern Workflow Solutions for Today’s Requirements</td>
<td>İltër Kalkancı</td>
<td>86</td>
</tr>
<tr>
<td>Linear Predictive Model for Audio Signals</td>
<td>Barış Erol, Bahri Çağlıyan, Sevgi Zübeyde Gürbüz</td>
<td>97</td>
</tr>
<tr>
<td>“Wallpaper Music” and Sonic Spices: Toward an Aesthetics of Sound Design in In the Mood for Love</td>
<td>Timmy Chen Chih-Ting</td>
<td>103</td>
</tr>
<tr>
<td>Indie Games: A Case Study</td>
<td>Massimo Avantaggiato</td>
<td>110</td>
</tr>
<tr>
<td>Do-It-Yourself Philosophy and Electronic Music Applications</td>
<td>Mustafa Karakaplan, Arda Eden, Halil İmik</td>
<td>120</td>
</tr>
<tr>
<td>Auralization as a Sound Quality Assessment Tool</td>
<td>Konca Şaher and Feridun Öziş</td>
<td>129</td>
</tr>
<tr>
<td>The Secret Theatre Revisited: Eavesdropping with Locative Audio</td>
<td>Pieter Verstraete</td>
<td>135</td>
</tr>
<tr>
<td>Audio Feature Extraction for Exploring Turkish Makam Music</td>
<td>Hasan Sercan Atlı, Burak Uyar, Sertan Şentürk, Barış Bozkurt, Xavier Serra</td>
<td>142</td>
</tr>
</tbody>
</table>
INTRODUCTION

ATMM (Audio Technologies for Music and Media) was a trilogy of international and interdisciplinary conferences (2012, 2013 and 2014) that focused on the various aspects of audio, audiovisual and music technologies for music and media, and, also, on the relationship between sound, music and image in both ‘traditional’ and ‘new’ media.

First of its kind in Turkey in the field of audio technologies, ATMM brought together professionals, academics, practitioners and students, coming from eleven different countries. In three conferences a total of 79 presentations, 6 workshops, 7 musical performances, 5 panels and 6 product demonstrations took place.
AUDIO TECHNOLOGIES FOR MUSIC AND MEDIA 2014
INTERNATIONAL CONFERENCE

ATMM 2014 Steering Committee
Ufuk Önen (Bilkent University)
Teoman Pasinlioğlu (Independent Researcher)
Mustafa Ertan (Interelektro)
Ahmet Gürata (Bilkent University)
Bülent Biryakoğlu (KV331 Audio)
Kürşat Pasinlioğlu (Voispectra)
Yusuf Akçura (Bilkent University)

ATMM 2014 Review Board
Barış Bozkurt (Bahçeşehir University)
John Krivit (The New England Institute of Art)
Tolga Tem (Sebit & METU)
Pieter Verstraete (Bilgi University & University of Exeter)

ATMM 2014 Assistant Members
Ipek Altun (Bilkent University)
Bilge Miraç Atcı (Bahçeşehir University)
Giray Bayer (Bilkent University)
Gercek Dorman (KV331 Audio)
Banu Şahin (Detmold University of Music)
Aycan Yücel (Bilkent University)

ATMM 2014 Assistants
İdil Acım
Ali Burak Parım
Günsu Çiftçi
Sezgin Akbudak
Ege Çakır
Feridun Gündeş
ATMM 2014 OPENING CEREMONY AND KEYNOTE

ATMM 2014 opened on Wednesday, November 12, at Bilkent University, Faculty of Art, Design and Architecture, with remarks by Bilkent University Rector Prof. Dr. Abdullah Atalar, Ufuk Önen and Teoman Pasinlioğlu.

Keynote speech of ATMM 2014, titled Rediscovering Creative Sound Engineering: a Challenge to Producers, Engineers, Artists, and Educators, was delivered by Pieter Snapper.

Snapper is the founder of the MIAM studios and co-founder of Babajim Istanbul Studios & Mastering, on contemporary opera production techniques. Snapper is a mastering engineer, producer, and composer of contemporary classical and electronic music. His works have been played in the U.S., Europe, and Asia. He has received awards from BMI, ASCAP, UC Berkeley, the Union League Foundation, and commissions from the Fromm Foundation at Harvard University, Yamaha Corporation of America, the ensemble Eighth Blackbird (three time Grammy-winners), and the Memphis Symphony Orchestra.
ATMM 2014 FEATURED PRESENTATIONS AND EVENTS

Barry Marshall (New England Institute of Art)
Strategies for Dealing with the Digital Disruptions of the Recording Industry
[12 November 2014 – 14:30-16:00 Ankara, FFB-022]
Workshop
13 November 2014 – 15:30-17:30 Ankara
14 November 2014 – 16:00-18:00 Beşiktaş, İstanbul

Tolga Tem (Sebit & Middle East Technical University)
The Art of Film Scoring
[12 November 2014 – 16:30-17:30 Ankara, FFB-022]

Ali Cenk Gedik (Dokuz Eylül University)
Audio Engineering as a Musical Performance: Towards an Ethnomusicology of Record Production
[13 November 2014 – 10:00-11:00 Ankara, FFB-005]

Cumhur Erkut (Aalborg University)
Sound and Music Computing: A Nordic Perspective
[13 November 2014 – 11:30-12:30 Ankara, FFB-022]

ATMM 2014 Hall
12-13 November 2014 – Ankara
Tolga Özoğlu (Roland Product Specialist)
Başar Ünder (Roland Product Specialist)
Artist
Doğaç Titiz (Roland Türkiye V-Drums Artist)
Şinasi Celehiroğlu (Demonstrator)
Can Tüzün (Roland Türkiye V-Drums Artist)
Gerçek Dorman (Roland Türkiye V-Drums)

Guitar & synthesizer: searching for new guitar sounds
Using electronic instruments in live performance, recording and mixing: AIRA, V-Drums and Handsonic
New directions in music technologies: SuperNATURAL, Analogue Circuit Behaviour (ACB), Multi Dimensional Processing (MDP)
Audio-Visual Technologies: Recording technologies and using MIDI in live video art Performances
Meet the Authors

Gerçek Dorman  Ufuk Önen
Arda Eden  Teoman Pasinlioğlu
Cihan Işıkhan  Kürşat Pasinlioğlu


Book signing hour with the authors Gerçek Dorman, Arda Eden, Cihan Işıkhan, Ufuk Önen, Teoman Pasinlioğlu and Kürşat Pasinlioğlu

Audio Tech Startup Forum

Beatografi
Cubic.fm
EasyGuitar
KV331
Seshouse

(13 November 2014 – 13:00-15:00 Ankara, FFB-022)

Audio Tech Startup Forum with Beatografi, Cubic.fm, EasyGuitar, KV331 and Seshouse.

Current and Future States of the Audio Industry in Turkey

Ender Balcı  Samim Mutluer (SF)
(Flat Müzik ve AudioCommando  Ufuk Önen (Bilkent University)
Community & Network)  Tolga Tolun (SAE Institute)
Görkem Çelikbilek (Asimetrik)  Çağan Tunali (Noiseist)
Gökhan Deneç (Babeshehir University)  Yılmaz Yeniyol (iPoint)
Cumhur Erkut (Aalborg University)

(14 November 2014 – 13:30-15:30 Beşiktaş, İstanbul)

Discussion of the current and the future states of the audio industry in Turkey from different perspectives including studio production, live sound, education, market and employment.
ATMM 2014 PAPER SESSIONS

SESSION A (12 NOVEMBER – 14:30-15:30 – ANKARA)

• Sparse Solutions in Audio Inpainting – Mehmet Erdal Özbek (İzmir University Electronics and Communications Engineering)
• A Survey Of Pitch Correction Methods – Zeki Hayran, Ali Burak Parım, Mehmet Saygıçn Seyfioğlu (TOBB University of Economics and Technology)

SESSION B (12 NOVEMBER – 16:00-17:30 – ANKARA)

• You Are a Target Market: Using News Media and Advertising to Create Generative Sound and Visual Art - Alayna Aine Hughes (Berklee College of Music, Valencia)
• Music and Audio Production for Documentary Films: Over the example of a Particular Documentary: Yılan Hikayeleri (Serpent Stories) – Gihan Işıkhan (Dokuz Eylül University)
• Theatrical Time Through Audio Technology: Steve Reich and Beryl Korot’s Three Tales – Jason Jedlicka (Indiana University)

SESSION C (13 NOVEMBER – 10:00-11:30 – ANKARA)

• Community Radio, Industrial Involvement: A Case Study – Patrick Quinn (Glasgow Caledonian University)
• Auralization as a Sound Quality Assessment Tool – Konca Şaher (Kadir Has University), Feridun Öziş (Dokuz Eylül University)
• Audio Feature Extraction for Exploring Turkish Makam Music – Hasan Sercan Atlı (Bahçeşehir University), Burak Uyar (Bahçeşehir University), Sertan Şentürk (Universitat Pompeu Fabra-Barcelona), Barış Bozkurt (Koç University), Xavier Serra (Universitat Pompeu Fabra-Barcelona)

SESSION D (13 NOVEMBER – 11:00-12:30 – ANKARA)

• Meniere’s Disease and the Making of Discord (Concerto for Meniere’s Disease and Electroacoustic Orchestra) – Robert Gross (University of North Texas), Kenneth Downey (Independent Researcher)
• Creative Use of Synthestration and MIDI Virtual Instruments in a Jazz Recording and Film Scoring Context – Daniel Walzer (University of Massachusetts-Lowell)
• Do-It-Yourself Philosophy and Electronic Music Applications – Mustafa Karakaplan, Arda Eden, Halil İmik (İnönü University)
SESSION E (13 NOVEMBER – 13:30-15:00 – ANKARA)

- Establishing a Signal Processing Methodology for Multitrack Mixing Education – C. Barkın Engin (Bahçeşehir University)
- The Current Challenges Facing the Sound Designers and Music Composers for Film, and Modern Workflow Solutions for Today’s Requirements – İltür Kalkancı (MMA – Modern Music Academy)
- Indie Games: A Case Study – Massimo Avantaggiato (Conservatorio G. Verdi Milano)

SESSION F (13 NOVEMBER – 15:30-17:30 – ANKARA)

- “Wallpaper Music” and Sonic Spices: Toward an Aesthetics of Sound Design in In the Mood for Love – Timmy Chen Chih-Ting (University of Hong Kong)
- A Transmedial Approach to the Distribution of Music in Online Environments: The Hype Machine – Buğü Melis Çağlayan (İstanbul Bilgi University)
- Messiaen, Composer of ‘Awe’: The Proto-Spectralist Pursuit of Divine Fear in Apparition de l’Église éternelle – Miriam Piilonen (Northwestern University)
- The Secret Theatre Revisited: Eavesdropping with Locative Audio – Pieter Verstraete (İstanbul Bilgi University Istanbul & University of Exeter)
Creative Use of Synthestration and MIDI Virtual Instruments in a Jazz Recording and Film Scoring Context

DANIEL A. WALZER
UNIVERSITY OF MASSACHUSETTS-LOWELL
walzerd@gmail.com

Abstract

This paper explores the musical and sonic possibilities of scoring a chamber orchestra piece to an existing jazz trio recording using MIDI virtual instruments commonly found in digital audio workstations and high-quality sample libraries. This presentation carefully examines MIDI orchestration as a creative process in film scoring and music production in a way that supports the visual narrative. Issues concerning tempo matching, authentic versus sampled instrumentation, scoring and arranging aesthetics, sonic balance and realism, manipulation of MIDI data, stereo imaging, performance authenticity and listening perspectives reflect compositional approaches towards a clear and balanced soundtrack where virtual instruments support the jazz recording and film score.

Original Project Background

“Sanctuary” is an instrumental composition for a chord-less jazz trio consisting of flugelhorn, acoustic bass and drums. Originally composed and recorded in 2007, the piece is reflective and somber and done as a jazz waltz. The melodic content is angular, with several notes reaching the upper register of the flugelhorn’s range. The melody is particularly challenging due to some octave leaps which evoke an aching sonority. “Sanctuary” is named partly for the uniqueness of the recording space; an old cathedral in Cincinnati, Ohio ultimately converted into an artist and studio community.

The cathedral’s studio space served as an ideal acoustical and creative environment with the tracking area measuring 18’x 40’ with ceilings nearly 30 feet high at the peak (Parmenter). Bruce and Jenny Bartlett observe that a proper acoustic setting positively accentuates the essential characteristics of a transparent jazz recording including cohesive balance, proper natural reverberation and accurate listener perspective (287). In this instance, the cathedral was the ideal space to capture an acoustic trio and provided inspiration for the musicians and engineer involved in the session.

The original "Sanctuary" arrangement featured a 16-bar vamp on a Dmin7 chord played by the acoustic bass and drummer with brushes. The trio played through the melody and chord change sketches after an initial set up and sound check. Recording in a relaxed yet open space provided ample opportunity for experimentation and last-minute creative decisions about solo order, melodic phrasing along with the rhythmic feel and flow of the piece. The studio’s desirable acoustics and spaciousness provided an exciting
catalyst to fully realize the recorded jazz composition (Eno, “Pro Session: The Studio as a Compositional Tool”). A finalized arrangement included the opening vamp followed by the melody, flugelhorn solo, then a drum solo played over the form of the song with walking acoustic bass. The trio recorded “Sanctuary” live in two separate takes without a click track.

**Tracking, Editing and Mixing**

The original recordings included many technical and creative decisions concerning placement of the musicians in the large room to help visibility and optimize signal flow. The subtle adjustments to microphone and musician placement ensured consistent sonic and performance quality of the recorded takes. Those meticulous decisions made the post-production mixing process fairly straightforward. Tzanetakis, Jones and McNally note distinct panning decisions of lead instruments of an acoustic jazz ensemble remain fairly consistent over time (4).

*Figure 1* Drums elevated on stage with overheads placed in an ORTF pattern
The strategic placement of the drums on the stage also compensated for the absence of an extra harmonic instrument. The trio’s interactive playing filled the necessary space. Spreading the flugelhorn and bass ten feet apart in a triangular shape reduced bleed between the instruments (Parmenter). The bass drum, snare and toms had close microphone placements with the overheads configured in an ORTF pattern after some experimentation.

![Image](image.png)

*Figure 2* A closer look at the drums and cathedral recording space.

The recording engineer carefully edited and mixed down “"Sanctuary"” a few weeks later to a stereo two-track master. Session edits included subtle pitch correction for a few of the melodic notes in the flugelhorn and a few short edits from different sections of the separate takes to form strong composite solos¹.

**Revisiting and Reworking**

While the "Sanctuary" sessions achieved desired results, the stereo master sat idle for more than five years. After completing other ensuing musical projects, a wonderful opportunity to revisit the song presented itself as a graduate school final thesis. After archiving the "Sanctuary" multi-track sessions, they were unfortunately lost due to a computer hard-drive crash a few years later. The remaining stereo mix from the original sessions survived garnering inspiration for a newly re-arranged piece for jazz trio and chamber orchestra using virtual instruments.

"Sanctuary" provided the musical canvas to sculpt a new robust accompaniment using virtual instruments in Pro Tools 10. The final project consisted of creating an original MIDI orchestration using techniques developed and refined in an Orchestration and Technology I class at Academy of Art University during the summer of 2013. In the following paragraphs I describe some potential sonic and musical challenges and outline some techniques used to create a new hybrid piece using MIDI instruments as a backdrop.

---

¹ “Sanctuary” was originally mixed by Matt Parmenter at the Ice Cream Factory Studio in Cincinnati, OH.
The Composing Dilemma

To reduce the potential sonic and musical issues at hand with hybrid workflows, the first compositional plan considered the key differences among sample libraries, acoustic instruments and recordings. Modern composers face a dilemma writing music that is playable by real musicians versus optimizing sample libraries to realize full creative scoring possibilities with technology (Asher, "Composing with Sample Libraries"). Fulfilling the scoring assignment within time and budgetary restraints required the use of high quality samples rather than hiring instrumentalists. Additionally, using Pro Tools as the primary digital audio workstation proved helpful in the mixing and future synchronization to video.

Initial planning of the orchestrations included experimentation of non-invasive timbrallayers, mapping the dynamic range of the original "Sanctuary" recording, and genre appropriateness. The original version of "Sanctuary" is best classified as jazz. The rhythmic feel of the melody, time pulse and improvisations swing throughout the original recording. Absolute labels are subjective, however the combination of swinging melodies, instrumental soloing, counterpoint and orchestral timbres places "Sanctuary" in the crossover realm of jazz and chamber music (Catalano, "Chamber Jazz"). Ultimately, the term "hybrid" classifies the entire project as a combination of traditional compositional principles aided by the competent use of technology.

Initial Challenges

The first stereo mix of "Sanctuary" captured a musical performance in a distinctly open recording space. New MIDI instrument tracks needed to conform and match the original recording tempo as closely as possible. The trio's performance was consistent but there were slight tempo fluctuations throughout. Fortunately, the Elastic Audio and Beat Detective capabilities found in Pro Tools allow for a relatively consistent tempo analysis and match through the meticulous creation of bar lines at key phrases in the audio track. Outlining a stable tempo map required some nudging of each marker, which ensured the transients were accurately represented without creating unwanted artifacts caused by time compression and expansion. The use of markers in Pro Tools to outline a detailed structure proved invaluable when orchestrating parts later along with session notes and comments.

Arrangement Structure

The "Sanctuary" arrangement is divided into five distinct sections including:

A) Bass and Drums Introduction (16 bars, full orchestration)
B) First Melodic Statement (Played twice for a total of 40 bars, melodic counterpoint in solo violin)
C) Flugelhorn Solo (One chorus, acoustic guitar comping followed by woodwind/string backgrounds)
D) Drum Solo (One chorus, pizzicato strings followed by woodwind figures the second time)
E) Final Melodic Statement (Played twice, counterpoint in brass and low strings, full held note)
Orchestrating new ideas required transcribing the original chord changes and melody to cultivate the most appropriate voicing options. Fortunately, the melody is quite high and provided enough space to use a variety of standard voicings including closed position, drop 2, and unison/octave pairings. Arranging decisions considered the absence of a chordal instrument on the original "Sanctuary" recording along with a close analysis of the walking bass lines outlining the harmony. Many of the new chord voicings came from experimentation on an acoustic piano. The original recording provided much harmonic freedom for the improvisers. Ultimately a more standardized approach proved useful in when creating arranged parts.

The technical aspects of Pro Tools control and manipulate audio transient information to match the original recording’s major characteristics through non-destructive editing. The project emphasized common techniques found in jazz arranging and orchestration including counterpoint, range, instrument combinations, key signatures, tempo, and the ability to tell a story through the composite “feel” of the MIDI ensemble. Aesthetically, a conscious effort identified the chamber orchestrations as supporting material, which highlighted the beautiful flugelhorn work of the original recording. As such, background and ensemble figures were not overly complicated and counter-melodies followed the same compositional approach.

In addition to the specific musical issues at hand, the sonic components of supporting MIDI orchestrations needed to match and compliment the distinctive acoustics of the original recording. The original tracks received little delay and reverberation during the mixing process, except the flugelhorn at subtle points during the solo sections.

**Sampled Instruments vs. Acoustic Tracks**

Pro Tools' Xpand!² provided the sounds for the orchestrations. Xpand!² is a robust software instrument that comes when purchasing the digital audio workstation. As a software synthesizer, it is capable of handling multiple timbres, which may be configured and routed to separate MIDI tracks, or layered for unique textures on a single instrument track. Frank Cook notes Xpand!² has an extensive sound library with thousands of individual patches and presets along with effects and advanced MIDI settings (121).

Matching individual patches in Xpand!² with the "Sanctuary" session required some careful planning. From a purely musical perspective, instrument combinations and layers needed to support the lead flugelhorn melody that compensated for the absent chordal accompaniment in the original recording. As a hybrid work, the opportunity existed to craft and layer distinct combinations of samples not commonly found in a standard chamber or studio orchestra score.
Xpand!² provides many flexible options with instrument panning, effects choices and routing. This helped integrate the layered samples into cohesive sets that filled the stereo image of the entire Pro Tools session. Multiple virtual instrument tracks enabled the brass, strings, and woodwinds to occupy their own separate place within the composite listening field. The original "Sanctuary" recordings had relatively conservative panning treatment, with primary focus on the drums. The acoustic tracks required no panning automation to keep the primary focus on instrument and microphone placement. "Sanctuary" received little alteration when imported into Pro Tools. As the centerpiece of the orchestration, all creative decisions revolved around supporting the original composition, while creating subtle changes to the depth of the entire piece.

The Tools and Initial Process

Once finalizing structural details of the recording, initial composing took place on a Mac Pro running Snow Leopard OS 10.6.8, Pro Tools 10 and an M-Audio Oxygen keyboard controller monitoring through Sony MD7506 headphones and various speaker configurations including Tannoy and Genelec for playback.

The Pro Tools session ran at 48/24 using the Mac's internal clock source. The "Sanctuary" file exported at 44.1/24 initially, then imported into the new session to begin orchestration. Perfectly emulating traditional placements of instruments was not the original intent of the supporting orchestration. Instrument sections were grouped based on timbral characteristics and blend within the Xpand!² platform.

Generally, Xpand!² was inserted as a plug-in on a series of instrument tracks. Creating four sound combinations through individual patches in Xpand!² proved more effective when considering the various sections in the chamber orchestra. The patch combinations were panned separately within the Xpand!² platform along with editing reverberation and other effects. Treating the composite sounds as "families" influenced the sequencing decisions positively. Matching sounds by timbral quality inspired more
legato keyboard playing in delicate passages. Lead instrument combinations required a slightly more aggressive treatment. Recording a melodic counterpoint, then background pads and finally solo section figures made the first orchestration pass fairly successful. Intimate familiarity with the original recording proved helpful along with Pro Tools markers, session notes and a written transcription of the original chords to outline the basic arrangement.

**Humanizing MIDI**

Each major section received some subtle MIDI editing through the Event Operations menu in Pro Tools. Recording ideas using the M-Audio controller involved quantization to the bar lines established through Beat Detective. In some areas, the information needed some manual adjustment to accurately simulate the tempo fluctuations contained in the "Sanctuary" master track.

The solo violin track served as a musical counterpoint to the flugelhorn melody. This track required some fine edits to note lengths in the MIDI Editor Window in Pro Tools. Through careful listening, solo violin note lengths needed to match the flugelhorn's phrasing without sounding forced. Even after originally quantizing to the bar line, some individual MIDI regions were moved to simulate anticipation of the chord changes and lock in with key accents in the cymbals. Using the drums as a timing center proved helpful in establishing some compromises in the rhythmic subdivisions. Ultimately this approach simulated a musical ensemble with slight phrasing differences among a group of players. Doubling the solo violin track with a mandolin-like sound effectively smoothed and filled out the counterpoint and required little added fine-tuning.

![Figure 4](image.png)

**Figure 4** Solo violin track supported with harp, banjo and mandolin.
The solo violin patch in Xpand!2 is ideal for legato passages due to its slower attack and slight delay. However, the timing needed to swing which required manual adjustment of individual MIDI regions on a 16\textsuperscript{th} note rhythmic grid in the MIDI Editor Window. The MIDI velocity changed slightly in each section to match the flugelhorn. Randomizing the velocity by 10\% also proved helpful to let the solo violin speak in the spaces when the flugelhorn does not play. Brass, woodwinds and string pads received similar editing treatments. Those tracks required less fine-tuning but their note lengths had to match each other musically, particularly on the opening 16-bar sequence before the melody starts.

During the single-chorus flugelhorn solo, an acoustic guitar patch propelled the harmonic accompaniment. Chord voicings were generally sparse, but the patch fit acoustically with the ambiance of the original recording. This track needed to sound as though it was responding to a jazz soloist in the moment. Initial tracking attempts were more rhythmically complex, but the final mix left more space for the flugelhorn to develop a lyrical solo. A Fender Rhodes patch doubled the acoustic guitar track one octave lower. Panning both tracks opposite each other filled out the space. In the second chorus, a small woodwind ensemble supports the acoustic guitar and flugelhorn with a gentle passage outlining the chord changes.

On the original recording the drums played one solo chorus while the bass player kept time and responded to accents. To fully support this section of the tune, pizzicato strings and multi-layered patches played short accents matching up with the bass drum and cymbals. This section of the arrangement was the most difficult to perfectly synchronize. The pizzicato sounds outlined the accents without seeming obtrusive. In this section, randomizing MIDI velocities approximately 25\% gave each sound a bit of space and authenticity. In a real context, the players would likely watch for visual cues from the drummer or perhaps a conductor. Rather than quantizing the regions to the point of non-recognition, musical compromise dictated that edits breathe in a fluid motion with the original recording. In some cases, moving individual regions slightly ahead and behind the beat simulated the rhythmic interpretation of classical musicians. In this case, simulating a live performance reflected subjective rhythmic interpretation of where the beat center resides.

![Figure 5](image)

**Figure 5** Pizzicato layers panned hard right during the drum solo.
After completing the MIDI orchestrations, routing effects via auxiliary sends made session navigation much easier. Individual tracks received equalization. Some instrument tracks required careful balancing to support the flugelhorn’s lyrical quality. The muted trumpet, lower reeds and supporting parts often received a modest boost at 300 Hz to reduce muddiness. Routing all instrument tracks to an auxiliary compressor with gentle limiting helped to glue the mix together. This approach saved room for additional processing. Most of the dynamics processing occurred during the initial keyboard performances along with optimizing the MIDI data to support the original recording. With effective MIDI controls and careful gain staging, aggressive dynamics processing proved ineffective.

Listening to the original jazz recording and MIDI orchestrations in solo mode served a useful purpose in drawing a complete sonic picture of the combined elements. While headphones revealed subtle nuances in the arrangement, playing back the finished session on multiple sound sources helped to address equalization and balance issues for later revision.

In keeping with genre appropriateness, sending all the Xpand!2 tracks to a common reverb effect accurately mirrored the original recording. Roey Izhaki observes that jazz listeners generally expect to hear a natural reverberation that accurately represents the original recorded performance (402).

The individual settings in Xpand!2 controlled reverberation for the majority of the MIDI orchestrations. Sending the instrument tracks to a composite reverb simply created a tasteful sense of space without distracting the listener from the original audio tracks.
The "Sanctuary" composite session received conservative dynamics processing and ambiance treatment. Panning decisions were open, and slightly more experimental. "Sanctuary" is a jazz piece. And while the focus remained to accurately represent the original recording, the MIDI instruments received additional stereo width enhancement to accentuate a reflective mood the piece inspires. Roey Izahaki explains the creative use of panorama can inspire distinct moods within recorded music (67). Panning the individual patches within Xpand!² combined with careful panning of individual tracks supported the recorded performance. Routing all instrument tracks to the stereo width enhancer plug-in not only supported the original recording, but also gave the composite MIDI orchestration a distinctive place within the mix without distracting the listener.

The final mix preparations included placing Maxim on the stereo master fader to optimize the output levels. Set with a fairly high threshold, the processor brought the final mix to an appropriate and comfortable listening volume. Testing the final mix on several monitoring systems dictated the need for basic mastering of the track. The final output left plenty of headroom with peak levels not exceeding -2 db. Maxim is versatile in its application and includes dithering before bouncing down the final session. Subtle adjustments to fader levels improved the quality of the final mix and proper digital gain staging removed any risk for peaks or distortion of the output signal.

**Film Use**

The finished composite mix served a wonderful purpose as the opening underscore for a short documentary film about the Harlem Renaissance painter Aaron Douglas. As part of a final graduate thesis project, the documentary included original music, sound design, interviews with artists familiar with Douglas’ creative output and scripted voice over narration.

The Aaron Douglas tribute focused primarily on his strong connections to Nashville, Fisk University and recent traveling exhibitions paying homage to his diverse paintings. Amy Kirschke notes little existing creative scholarship regarding the powerful impact of Aaron Douglas’ influence during the Harlem Renaissance and beyond (xiii).
"Sanctuary", as a fully realized hybrid piece, supported the opening credits through appropriate pacing, meter and sonic choice. Approximately 90 seconds of the original track underscored the first series of images. The openness and moodiness of "Sanctuary" foreshadowed the multifaceted themes featured in Douglas' paintings. Most importantly, "Sanctuary" provided the pace in which the video edits transitioned from one sequence to another. The relaxed swing feel of the original recording directly influenced a gradual transition through the opening sequence, refraining from jump cuts and abrupt transitions. The pace of the opening sequence closely matched the overall musical tempo of "Sanctuary".

MIDI optimization techniques refined during the "Sanctuary" orchestration sessions were fully realized throughout the rest of the Douglas film project. Several of the film's scores utilized a combination of acoustic instruments and software-based production methods. These methods include randomizing note velocities, use of multi-layered patches, creative routing options, and automating MIDI and audio parameters during the final mix. While the vast majority of the sounds in the documentary came from sample libraries, they tastefully support the visual image while providing a reasonable sense of performance authenticity. These optimization techniques transition effectively to other sample libraries and digital audio workstations. Budgetary restrictions limited the scoring to mostly virtual instruments. While challenging, manipulating the MIDI data proved effective to create the musical elements in the film.

Pro Tools served as the primary mixing platform for the sound design, effects, narration and music in the Douglas tribute. The "Sanctuary" underscore received little post-production treatment. The track starts the documentary with no additional dialog or sound design elements. Intentionally placing "Sanctuary" at the very beginning established a reflective yet atmospheric mood for the rest of the underscore. Production engineers mixed and locked all effects, music, and narration to picture in March 2014.

Orchestrating new parts to an existing audio recording presents a number of musical, technical and aesthetic challenges. Through careful and meticulous session organization, musical sensitivity, and thoughtful experimentation, it is possible to blend live audio recordings with virtual instruments. There is a danger in over-scoring a piece simply because the technology exists. Just as traditional arrangers, orchestrators and composers develop ideas through analysis and refinement, virtual instruments should be considered one of many tools available to fully realize a new score. No single digital audio workstation, software instrument or sample library provides every sound a composer might need. However, most platforms utilize a set of common tools for adjusting velocity, panorama, mix automation, filters, quantization and synthesizer optimization. When creatively used, these tools can enhance existing recordings in multiple genres.

Successful implementation of MIDI virtual instruments requires a practical focus on the audio mix, sonic realism and composer's intention. "Sanctuary" received conservative treatment due to its subtle and open jazz feel. Adjusting MIDI performance information does not replace real players on real instruments. New media composers optimizing MIDI parameters should thoughtfully reference classic recordings in the style they choose to emulate through virtual orchestration and arranging. Underscoring visual formats may consider a variety of individual and blended approaches that emphasize shared musical ideas, which tastefully compliment each other and entice the listener. As
sample libraries continually improve in sonic quality and cost effectiveness, they present exciting possibilities in future compositional projects.

REFERENCES


Parmenter, M. Personal interview. 11 September 2014.


2 Special thanks to John Freyermuth for extensive post-production mix assistance on the Aaron Douglas documentary.
Sparse Solutions in Audio Inpainting

MEHMET ERDAL ÖZBEK
İZMIR UNIVERSITY ELECTRONICS AND COMMUNICATIONS ENGINEERING
erdal.ozbek@izmir.edu.tr

Abstract

Audio inpainting describes the restoration process in audio signals. The degradations in the audio signal can be removed using many signal processing techniques and are replaced with the recovered part of the signal. Recent solutions rely on the sparsity of the signal based on the assumption that the signals can be represented with a few number of non-zero coefficients when they expanded in a suitable basis. Using this overcomplete representation, signals are approximated by a sparse linear combination of elementary waveforms forming a dictionary. The approximated signal is obtained by finding the smallest set of atoms from the dictionary via solving the optimization problem. This study examines recent advances in sparse solutions which demonstrate improvements in the signal-to-noise ratio compared to the conventional approaches.

Introduction

Audio restoration is the process of removing the degradations in the sound recordings. These degradations can be errors like hiss, click, and noise which are heard but are irrelevant to the sound recordings; they may occur because of the limitations due to recording, transfer, and broadcasting systems as in clipping; or due to unavailability of the signal such that some part of the signal is called as missing. For example, cleaning sound events like mobile phone ringing or coughing in a musical recording, filling the empty sections after cleaning, and restoring the record such that no processing done could be observed, or removal of artefacts while renovation and digitalization of old sound or music recordings, constitutes the aim of restoration. Orcalli has investigated the fundamental principles and methodologies of the audio restoration and its technical and scientific evolution (307).

There are different types of degradation as stated in Godsill and Rayner, therefore there have been many techniques to recover satisfactory representation of data using the information extracted from its available (undegraded or clean) parts (6). These include signal processing techniques based on autoregressive modelling, interpolation and extrapolation, or statistical modelling. One of the early studies in audio restoration uses adaptive interpolation and performs very efficiently (Janssen et al. 317). As the targeted degradations might be different, the proposed solutions are likely to follow diverse terminology depending on the context. Nevertheless, the aim in recovery/reconstruction/restoration is to find/estimate and then replace/impute the unreliable/unavailable/missing data with the estimated/restored data. The overall
framework is termed as *audio inpainting* based on the inpainting term describing the restoration in image paintings.

In our work, particular interest is given to the restoration of the clipped and missing audio where some part of the audio signal is not available, or its amplitude is clipped to some certain level. When the audio signal exceeds the maximum allowed amplitude level of the system, it is limited with this value. Figure 1 displays an example signal where the amplitude is clipped to a fixed value. The original signal can be restored by estimating the signal at the clipped regions and replacing it. Inherently, the clipped samples are arranged in groups and their locations are not randomly distributed but determined by the varying amplitudes of the signal and the clipping level. Clipping also introduces nonlinear distortion by modifying existing frequencies or adding new ones (harmonics). Thus, there is no pre-defined general solution proposal for de-clipping as in noise filtering. Accordingly, Godsill and Rayner separate degradations as either local or global (6). In localised degradations only certain samples of the waveform is affected. Global degradations affect all samples of the waveform. On the other hand, missing parts may be encountered at any time with variable durations throughout the signal. While shorter missing intervals may occur as in the removal of clicks, longer data may be lost during the transmission of packets. Any general information extracted from the whole signal (such as the noise level) or any available a priori information can be used for recovery, however the temporal index of the clipped/missing data is mainly of concern for audio inpainting.

One of the recent solutions to the audio inpainting is based on sparse representations. Sparsity depends on the assumption that the signals can be represented with a few number of non-zero coefficients when they expanded in a suitable basis. Using this overcomplete representation, signals can be approximated by a sparse linear combination of elementary waveforms forming a dictionary.

This overcomplete dictionary can be built by either using a pre-defined set of functions, or designing it such that it adapts itself for a given set of signal examples. In audio signal processing the most familiar representation is the Fourier representation (discrete Fourier transform or DFT). By performing a DFT analysis, it can be easily observed that the audio signals have energies in certain frequency bins where in most of the frequencies the energy is relatively low or zero, thus sparse. In other words, most
signals are not exactly or perfectly sparse but rather approximately sparse. Other invertible transforms such as the discrete cosine transform (DCT), or the discrete wavelet transform (DWT) are appealing because they are simple and easy to calculate with fast algorithms (Plumbley 995).

Based on the framework defined in Adler et al., one can consider the audio data which has the partitions composed of degraded data part and reliable data part based on the temporal index (922). Then, the audio inpainting problem is defined as the recovery of the audio data coefficients based on the knowledge of the available information about the data and the corresponding partition information.

Given the signal and the selected dictionary, if the signal is (approximately) sparse in some transform domain, the signal can be reconstructed by using only few elements (atoms) from the dictionary. Then the aim is to find the smallest set of atoms from the dictionary to represent it. The natural measure of sparsity is obtained by counting the number of non-zero coefficients via $\ell_0$ norm. Then, one may recover an estimate of the signal by solving the constrained $\ell_0$ norm minimization problem for a given approximation error threshold.

However, $\ell_0$ norm minimization is a combinatorial problem and determination of an exact sparse representation is computationally infeasible (Davis, Mallat, and Avellaneda 61). Therefore, computationally more tractable ways have been proposed by relaxing the $\ell_0$ norm to $\ell_1$ norm as in basis pursuit (BP) and greedy methods, as in matching pursuit, orthogonal matching pursuit (OMP), and iterative thresholding algorithms. Norm relaxation converts the problem to be a convex optimization where both BP and OMP are successful in representing a sparse signal using a wide class of dictionaries (DCT, Gabor, etc.) including unions of dictionaries, or adaptive learning dictionaries.

The method of Adler et al. uses the OMP algorithm by adding the clipping and missing data information as constraints to the optimization problem. Therefore, degraded audio signals were restored using the information obtained from the remaining clean part. The performance of the inpainting algorithms are then evaluated and compared considering the original signal and its estimate using the signal-to-noise ratio (SNR) measure computed for variable levels of artificially clipped samples and missing durations of speech and music signals.

Alternative approaches use re-weighted $\ell_1$ norm solutions where the norm is weighted with weight coefficients that are iteratively updated (Weinstein and Wakin 4). As re-weighted $\ell_1$ norm solutions approach to $\ell_0$ norm solutions better than $\ell_1$ norm, they provide better performance than BP with clipping constraints. Similar to the constrained OMP algorithm, when the information coming from the clipping index is included in the optimization problem of sparsity, SNR performance increases.

Other recent approaches include social sparsity which incorporates the temporal correlation of the Gabor dictionary coefficients as a priori information (Siedenburg 1577). Secondly, by adding the constraint that the reconstructed samples must be equal or greater (in absolute value) than the clipping level to the de-clipping problem, convex problem is formulated as an unconstrained convex relaxation.

Remarkably, audio inpainting performance evaluations have been performed not only with objective SNR measure but also with subjective evaluations through listening tests (Defraene 2634). As the perceived audio quality may not be necessarily coincide with the signal restoration quality, additional information coming from the human auditory perception may still be incorporated to the optimization problem.
Conclusion
In this article a brief overview of sparse solutions in audio inpainting is given. In the general framework of compressed sensing, it is seen that general sparse solutions and pursuit algorithms can be applied to restore the degraded part of the audio signals. As demonstrated in the works that use constrained algorithms, the inclusion of the clipping constraints has been resulted with performance improvement. Adding side information as in the case of perception shows that there is still room for possible enhancements. Although not discussed, the computer load for obtaining an optimum solution is relatively high even for the fast algorithms. Our studies are continuing to achieve better and faster solutions.

Acknowledgements
This work has been initiated based on the support of TUBITAK 2219 International Post-Doctoral Research Fellowship Programme conducted with Axel Röbel at Institut de Recherche et Coordination Acoustique/Musique (IRCAM).

REFERENCES


Precis for Celtic Music Radio (CMR) Final

PATRICK QUINN
GLASGOW CALEDONIAN UNIVERSITY, GLASGOW, SCOTLAND, UNITED KINGDOM
P.Quinn@gcu.ac.uk

Introduction

Industrial involvement is key to the success of vocationally oriented degree programmes. In audio, although many employers can be generous with time and involvement, the actual amount of "hands on" time can often be limited for various reasons including Health and Safety issues or time pressures. Another route to industrial involvement is through Social Enterprises - organisations that although commercial in nature are designed to be of value to the community. One area of social enterprise, community radio, has been set up to serve local communities in the UK by providing a localised radio experience. Regulated by the UK’s independent regulation authority OFCOM, they typically broadcast locally within a radius of a few kilometres using low powered transmitters and globally over the Internet.

This paper describes the involvement of audio students at Glasgow Caledonian University (GCU) with the community radio station Celtic Music Radio (CMR). CMR is different from typical community radio stations in the UK in that it primarily serves a "community of interest" around the world, namely those interested in Celtic/Roots music and serves as an outlet for emerging artists from Scotland and beyond.

The relationship between GCU and CMR now provides students with an opportunity to work on a live radio station both in a studio setting and in an outside broadcasting environment as part of the world’s largest Celtic music festival, Celtic Connections. Students have clearly benefited by being able to gain experience under the direction of station operational staff in a "real world" environment freed from the undue pressures and constraints of more commercial broadcasters. CMR, like many community radios has limited resources, so benefits by being able be more ambitious in its programming and by being able to provide a better quality of output to its listeners.

2 Audio Technology at GCU

Glasgow Caledonian University has been offering degrees in Audio Technology since 1994. Initially the degree on offer was only the BSc(Hons) in Music Technology with Electronics but after an evaluation in 2001 a suite of degrees programmes in Audio Technology was developed which were broader in outlook and encompassed a wide range of topics involving audio including signal processing, acoustics, live sound, studio production, electronics amongst other topics. This suite currently involves two degrees, namely the BSc(Hons) Audio Technology with Electronics and the BSc(Hons) in Audio Technology with Multimedia. Both programmes offer an in depth coverage of the technology behind audio systems emphasising the underpinning theoretical concepts but as can be seen from the curriculum given on the University’s website, each equally has a focus on developing good practical skills, which often can be essential for gaining that first position in employment in the audio
3 Industrial Involvement

GCU, in common with many vocationally focussed Universities, includes in its mission statement the clear intent to work in partnership with business to provide an "outstanding inclusive learning environment". To this end the teaching team for Audio Technology have always included industrial involvement as part of teaching and learning. This helps to ensure that as Scheirman highlights, the curriculum should remain relevant to industry and that links between education and industry are enhanced. This involvement has included guest lectures from many areas of audio including acoustics, broadcasting, equipment manufacturing, sound design and audio production and the teaching team have always found representatives from industry keen to participate. In addition site visits have played an important part in the curriculum and of late have included trips to recording studios, concert halls, research facilities, broadcast studios etc. More recently the teaching team have developed a number of links that have allowed students to gain work experience in a variety of organisation in various areas of audio, including broadcasting, live sound, audio visual hire, recording studios and audio systems design and sales.

Any involvement with industry is beneficial to students (and often staff too) in terms of broadening their exposure to industrial practises, gaining a better understanding of the requirements of the workplace in both technical and in more general terms and in many cases being able to hear at first hand about the diverse experiences of those that already work in audio. These all apply directly to the case of work experience however the ability of students to actually gain additional practical skills does depend on the nature of the work experience organisation. In the experience of the author, there are some organisations, where a variety of circumstances including Health and Safety or time/commercial pressures can limit a students' exposure to "hand on" activities and therefore students can spend a significant component of their time shadowing existing employees.

In this respect Social Enterprises that operate in the area of audio are an ideal source of work experience opportunities. By their very nature social enterprises exist to provide a service without the more overt commercial pressures that shareholders that might dictate. One area of social enterprise that fits this model is that of community radio.

4 Community Radio and Celtic Music Radio

Community radio as the title suggests exist to serve a local community by providing programming that is not available by mainstream commercial media (Bosch 426). Community radio has a background in facilitating opportunities for alternative viewpoints to be heard, typically in environments with very strict control of other forms of radio and media in general.

Community radio in the UK has perhaps been developed with a less of a political focus but with one that serves diverse groups such as particular ethnic groups or those with a common interest. Regulated by OFCOM, licenses are granted to community radio stations for five years at a time and stations are monitored to ensure that they conform to the requirements of the license and meet the aims of the initial bid.

Celtic Music Radio (CMR) was formed in 2004 to provide a platform for Celtic Music.
Inspired by the world's largest Celtic Music festival, Celtic Connections, which is held for between 2 and 3 weeks each January in Glasgow, Scotland, the founders recognised that this genre of music and indeed this important element of Scottish and Celtic culture was not getting the exposure that it merited. Although a healthy live scene of clubs and festivals existed (and has indeed grown) the opportunities for artists to be heard on radio was limited. Following on from a renewed interest in Scotland's Culture and cultural infrastructure after the re-introduction of Scotland's Parliament in 1999, CMR was set up to provide a suitable opportunity for this genre of music to be heard and therefore support Celtic/roots music within Scotland and beyond. In addition CMR incorporated training as a secondary objective within its remit.

Initially only broadcasting for the duration of Celtic Connections on a Restricted Service License, CMR was awarded a full community license which allowed it to broadcast all year round on Medium Wave (MW)/AM which started on January 2008. After six successful years of broadcasting on MW and online, and a further renewal of its license, CMR moved to broadcasting on FM in July 2014. This has allowed it to reach a wider local audience and to grow its diverse and ambitious range of programming including live guests, franchised programmes and outside broadcasts from various festivals around Scotland and Ireland. As is the case with many community radio stations, CMR runs with a very small budget and has received plaudits for its ambitions in terms of its programming given its funding limitations.

Audio students from GCU first got involved with CMR at Celtic Connections 2012. One of the highlights of Celtic Connections is the Danny Kyle Open Stage. Originally hosted by Danny Kyle, a keen promoter of Scottish culture, it was at first an opportunity for anyone to turn up and perform to festival-goers in the tradition of the informal music session. As the festival grew the open stage quickly developed into a showcase of the best new talent in Celtic and Roots music. Artists had soon to apply before hand to take part and are now pre-selected to perform at what remains a free but ticketed daily event. At the end of each festival the six best acts are chosen and rewarded with a gig on the final night of the festival in front of a paying audience, plus the opportunity of a support slot at the following year’s festival.
For the duration of Celtic Connections, CMR presents a live outside broadcast programme from the venue of the Danny Kyle Open Stage. This daily live programme, which starts at 10 a.m, culminates in the performances from the open stage from 5 until 7 p.m. Many of these artists will have not played in public before let alone be broadcast live worldwide. The first 7 hours of the programme consists of interviews and live performances of artists who are involved in the main festival and also the many artists who are on the fringes of the festival but involved in Celtic/Roots music making across the world. These guest slots are interspersed with a revised playlist of material from CMR’s usual output.

As mentioned previously CMR runs with a very small budget and has the ability as such to run a simple OB, but lacks the equipment and staffing to cater fully for its ambitions at Celtic Connections. So in 2012 Audio Technology students were first involved in assisting with broadcasting by providing and operating some additional equipment in the form of a mixing desk and microphones which allowed a much improved quality of output for the range of guests who were performing at the OB. From the involvement of 2 students for around 5 days in total in 2012, GCU’s involvement has grown in 2014 to providing all the audio facilities (with the exception of the playback systems and network connection) for the OB with a crew of 15 students working in teams for 10 hours each day over 14 days. Due to the expansion of the OB in 2014 and because the students involved came from all 4 years of the Audio Technology degree and had varying levels of experience in practical scenarios, two days of training were carried out involving a professional broadcast sound supervisor and one of the directors/presenters of CMR. These sessions proved to be highly valuable in contextualising the students’ understanding of the CMR setup at Celtic Connections and the expected demands placed on students.

The increased size of GCU’s involvement also led students having to work in teams and take on various defined roles as part of the OB. These roles were

- Broadcast assistant, where they operated the broadcast mixing desk, cued up CDs and liaised between the presenters and the other technical crew.
- Music Engineer, where they operated a separate mixing desk for any guests who were going to perform live. The guests ranged from singer/songwriters to 6/7 piece bands with a range of acoustic and electric instruments.
- Sound assistant, who assisted the Music Engineer in the setup of their system for live mixing.

Looking ahead, it is planned for GCU and CMR to work even more closely in the future and discussions have already taken place which could lead to:

- GCU providing enhanced OB facilities at other


6 Feedback from CMR and Students

Nyquist Theorem, also known as the Sampling Theorem, states that if any continuous signal with the highest frequency $f$ is sampled with a sampling rate greater than or equal to $2f$ then it can be reconstructed completely from its samples. In this relation, the highest frequency allowed in the signal, which is $f$, is called the Nyquist frequency and the corresponding double value $2f$ is the Nyquist Rate or Sampling Rate. A typical frequency distribution for a continuous signal is given in Figure-1 below.

In preparation for this paper the author interviewed Ross McFadyen, one of the directors/presenters of Celtic Music Radio. The overall view from the station is that the involvement of GCU and its students allows CMR to produce a significantly higher quality of output for the duration of Celtic Connections than would have been possible otherwise. As a community radio station CMR has ambitions, which are in many ways over-stretched by its current resources, a gap which GCU is able to bridge successfully and to the benefit of both parties. He commented further that an added advantage of GCU’s involvement is that it eases the burden on the volunteer presenting staff of CMR who can concentrate on the production of their programming rather than the technical issues associated with broadcasting from a location using equipment that they are unfamiliar with. As mentioned previously training is an integral part of CMR’s remit and that extends to its own volunteers as much as with external partners such as GCU.

To gain the students point of view a questionnaire was administered online. Their views on the experience were overwhelmingly positive with all respondents stating that they felt very or reasonably confident in working in any of the three roles identified above in the future as a result of their experiences on the OB. Given that a large minority had no experience in the broadcast assistant and sound assistant roles, that appears to be a very positive result.

The introduction of specific training for 2014 was highly valued by the students and they stated that they felt well prepared for working at the OB. The students listed a wide range of other high points and benefits from being involved including:

- Real world experience
- Working in teams
- Knowing that your work is being listened to worldwide
- Live mixing opportunities

There were a small number of suggestions on ways to improve the experience including allowing students to mix the live bands for the open stage for broadcast every evening but all of those that responded rated the experience as being excellent or very good.

7 Conclusions

It is clear that industrial involvement is of great value to students on an audio focussed degree programme. In many ways the greatest benefit comes from increased opportunities to have hands on practical experiences, which often can only be achieved in organisations with less commercial/time pressures such as community radio stations. In this case study students have gained valuable practical experience working on live radio in a pressurised situation, but being a community radio station mistakes are of far less consequence than for a commercial environment and the benefits of helping CMR meet its ambitions far outweigh any negatives.
OBs for CMR throughout Scotland and Ireland
• GCU providing programming for CMR by recording many of the local concerts that take place within Glasgow and its environs
• GCU and CMR combining their experience and collaborating in providing training to other community radio stations across Scotland

REFERENCES


A Transmedial Approach to the Distribution of Music in Online Environments: The Hype Machine

BUĞU MELİS ÇAĞLAYAN

ISTANBUL BILGI UNIVERSITY COMMUNICATION SCIENCES
bugumeliscaglayan@gmail.com

Introduction

A. A New Dawn for Music
Marshall McLuhan once stated the power of phonetics, thus the mere act of hearing, is undefeated and using sound as a tool of systematical communication is a transforming experience. (118) Jacques Attali agreed, underlining that the postmodern society no longer needs to see or feel, people only need to hear to perceive their own reality, and music is the most transcendental way to do this. Music industry, thus comes with great importance, and in the Webified culture of today, a structural shift can be observed. The industry, as well as the experience and production as whole, has been affected with various technological improvements, particularly the Internet. So if one would think that music, and even sound from a greater perspective, has always evolved hand in hand with technology -the transition from tapes to CDs and iPods replacing your once-beloved Disc Man; this comes by no surprise. Fast forwarding to 21st century, today marks an ever-burgeoning era for music where it is impossible even to mention such linear shifts.

Previous decades have witnessed a major and rather drastic change for music industry, especially for everyday music listening habits. Radio illustrates a perfect example; which was widely accepted as the main distribution and consumption platform for music until 70s. Decades after, the audience has been introduced with podcasts, radio-on-demands and various other digital streaming platforms; which can still be regarded as a rousing yet slightly exhausting experience. The fact that radio, as a medium, has always embraced various cross-media practices raised the bar for what it will transform into, yet the evolution of the radio (from terrestrial to online) failed to meet the expectations because of its rather linear progress, which eventually led to the rise of online streamers, offering a more personalized and interactive experience without the shock effect. Founded by Anthony Volodkin, a then-computer sciences student, in 2005 as an online music blog aggregator then becoming an independent music platform encompassing both online and offline environments, Hype Machine manages to create a transmedial narrative for music in the depths of Internet Galaxy.

When it comes to examining music as a new media product, the literature is still fairly limited: Studies analyzing bands as transmedia products are quite common, yet the online platforms which distribute and even affect the production tendencies of music have been tossed out when it comes to extending a transmedia-oriented dialogue. As Henry Jenkins highlights, "most often, transmedia stories are based not on individual characters [...] but rather complex fictional worlds which can sustain multiple interrelated characters and their stories." Although this might be confusing, partially because talking about characters and stories while discussing music industry can be regarded as misleading, if
not rather ‘fictional’, the individualistic boundaries must be overcome to truly understand the transmedial characteristics of music: The core aim of this study lies beneath the power of adapting and evolving ideas and frameworks on transmedia storytelling to the current consumption and distribution narratives of music in online environments by investigates the encompassing, dispersed and equally converged transmedia experience provided by Hype Machine, which rejects the horizontally integrated structure of traditional radio and spreads to various platforms with tailor-cut experiences.

B. Theory and Methodology

This article mainly revolves around Henry Jenkins’ ideas on media convergence and transmedia. Along with his online articles discussing further on the definition and boundaries of transmedia, his well-known book Convergence Culture: Where Old and New Media Collide (2006) serves as a major theoretical reference to discuss the case of Hype Machine. The conceptual framework will be supported by Lev Manovich’s book The Language of New Media (2001). The core challenge, however, is particularly based on the lack of literature on the specific field; as sound, music and distribution mediums including radio do not particularly attract the attention of new media theorists. Resulting in an inevitable scarcity on related researches, music is often underestimated as a transmedia product in a micro perspective; as Jenkins exemplifies with Nine Inch Nail’s ARG efforts or hip-hop’s transmedia-related characteristics as a genre and culture. What makes this view problematic is that music is now connotated with various broader experiences instead of a mere aural or cultural one -there are innumerable echoes forming a greater picture; along with the conceptual ambiguity between transmedia storytelling and transmedia branding.

The methodology is based on the critical interpretation of aforementioned literature on new media as well as interviews with Hype Machine’s founder Anthony Volodkin. Previous surveys, data and demographics on the field -especially on the distribution and consumption of music in online environments- will be examined alongside.

2. Another Future for Music in Online Environments

A. How Does Transmedia Sound?

21st Century has thoroughly affected the way people interact with music. Once strictly limited to physical releases, radio & TV broadcasts and live performances, today marks an immense habitat for accessing music -mostly resulting from digital methods streaming innumerable songs in various formats. (Krause et al 1) A 2013-study reveals that among the younger audience, music is mostly played via digital formats -including MP3 players, computers, smart phones and tablet devices. Although the same study reveals that the older individuals stick to the habits by still using CD players or opting for the radio, both facts combined give a clear idea on the future trends of music industry. (Krause et al 9)

Whilst perceiving online environment as whole, including any music-related experiences, it has to be kept in mind that Internet has transformed into a super platform embracing various mediums instead of being a mere media itself. The reason behind this misconception is particularly based on the ambiguity with previous research "suggesting
no clear answer to the question of whether the Internet is a substitute for or a complement to traditional media outlets”. (Gaskins & Jerit 190) The idea of replacing the traditional with the new results in a crisis while interpreting their messages, yet transmedia can as well be helpful to overcome this vain tendency of seeing the internet-related shift dangerously linear.

According to Jenkins, "transmedia storytelling represents a process where integral elements of a fiction get dispersed systematically across multiple delivery channels for the purpose of creating a unified and coordinated entertainment experience". (1) He expands the dialogue by mentioning a term coined by Pierre Lévy, by highlighting that transmedia storytelling reflects the ideal aesthetics of the "collective intelligence" era. What matters most in this definition, is not the ideal aesthetics, but what collective intelligence connotes instead. As a French philosopher focusing on the theoretical implications of cyberspace and digital communication, Lévy expresses that art in this collective intelligence era functions in a different way; interpreting any creative and artistic effort as "a cultural attractor, drawing together like-minded individuals to form new knowledge communities". (Jenkins 4) This is rather significant, as communities can as well be understood as a delivery channel among themselves, where the online narrative gets dispersed and elaborated. Keep the previous sentence in mind, as this might be a good starting point to throttle the conversation; but first, let’s talk a bit about Jenkins’ ideas on what transmedia storytelling is.

With the light of the aforementioned definition, it can be said that every medium included in the chain makes a contribution to the overall narrative. Needless to say, each contribution is unique, as all mediums are different by nature; ending up in a media synergy with different extensions (varying from creating enthusiasm and engagement to introducing complimentary side narratives), segmentations (addressing divergent audience groups with each medium) and performative dimensions (particularly related with the idea of community-oriented fan fiction). (Jenkins 1) The environment that gave birth to transmedia is equally as important as the definitions. As the first new media scholar officially welcoming us to the convergence culture, Jenkins came up with the idea of old and new media colliding, coming together and reinventing each other along with the audience. With Jenkins’ own words, convergence is "the flow of content across multiple media platforms, the cooperation between multiple media industries, and the migratory behavior of media audiences who will go almost anywhere in search of the kind of entertainment they want". (Jenkins 2) American scholar Jay David Bolter’s concept of remediation deserves to be mentioned now, because similar to convergence, remediation also "contrasts with the digital revolution model which assumed old media would be replaced by the new media". (Jenkins 1)

According to Jenkins, not all the acts of remediation should necessarily be treated as forms of transmedia, though. Still, it is an apparent fact that remediation is a core concept for understanding the broader territory. Lev Manovich’s definition of remediation is "the practice of putting together a media object from already existing commercially distributed media elements existed with old media, but new media technology further standardized it and made it much easier to perform" (130) For the particular case of The Hype Machine, a rather representational form of remediation can be discussed. But before getting further on our case, a brief look at the history of modern music industry is essential. For music, remediation is an indispensable part of evolving. Be it from a consuming perspective, and be it for recording, the endless act of remediation has always been a vicious loop; the invention of radio and the transition to television, cassettes attempting
to replace vinyls and CD's eventually wiping both out... Not a single technological leap has completely been independent from the prior one, they can coexist as much as they can until the older media slowly fades away -mainly to the (non-mainstream) kitsch rank. Therefore, a new concept of representational remediation can at least be considered; however similar it may be to the logic of hypermediacy, their partial incompossibility of (physically) coexisting leads to this conclusion. Nonetheless, with an aim to explain the dissemination of content, what Manovich points out can as well justify the emergence of Peer to Peer (P2P) platforms, along with music blogs constantly reinventing the dialogue between music and new media: "Rather than assembling more media recordings of reality, culture is now busy reworking, recombining, and analyzing already accumulated media material" (131).

With an attempt to illustrate a subtle embodiment of convergence in online environments -and in this case, remediation, Simone Murray elucidates; "employing the term 'content streaming' to describe the migration of content from one platform to another expands the term's strict meaning within media industries, namely the delivery of video or audio content via the internet". (9) The migration of content may not necessarily refer to a reciprocal dialogue between platforms all times; yet a connection does exist. The outdated interpretation of the process merely as one medium replacing the other is an arrogant, if not a completely inaccurate perspective. Let's rewind for a second to Lévy's thoughts on the era of collective intelligence, where the like-minded individuals form knowledge communities with the help of cultural attractors, in our case music. Participatory culture and interactivity are thus two terms to be focused on, both of which can as well be thought as signifiers of transmedia generation, convergence culture and collective intelligence. In the age of digital convergence, interactivity is perceived as de facto; after all, this is how World Wide Web works. Participatory culture, on the other hand, can both be supported by and enhance interactivity; offering globally accessible ways of expression which might pass beyond the limits of original narrative. Taking these as a starting point, Jenkins' seven principles of transmedia storytelling can be helpful to decode Hype Machine’s success as a transmedia experience.

B. The Future of All Media: Hype Machine

In a 2007-article on CNN Money, John Heilemann searches for the essence of Hype Machine -which "could define the future of more than just music", as he claims. In his article, Heilemann suggests "with a little imagination and daring, the Hype Machine (or something like it) could evolve into the single indispensable site in the Webified world of music". Gawker Media founder Nick Denton’s statement is far more daring though, when he said Hype Machine is "the future of all media". (1, 2) If all these sound too exaggerated, a brief look at the history of Hype Machine might be helpful. As an MP3 blog aggregator website, Hype Machine was created by computer science student Anthony Volodkin in New York, 2005. The idea was actually born out of various personal experiences of Volodkin, through which he realized music magazines and radio broadcasts were not sufficient to find new music. "At this time", Volodkin says while referring to the time when he first came up with the idea of Hype Machine,

the concept of 'music blogs' was very new -most people hadn't heard of them [...] It was amazing, because a few years before, the whole Napster thing went down, and pretty much nobody did anything online after that -it was just kind of dead. I started feeling like there must be some way for me to have a better look at what was going on there. (Rapp 2)
The system of Hype Machine operates as following: Users can listen to Hype Machine’s streams as a visitor, but the consequent amount of the music they can stream has a daily limit where they are reminded to sign up for a free membership; after all, that’s when the fun starts. Once signed-up, a user can stream various new music, create their own favorite list, discover remixes and peek at their friends’ everyday music listening habits. The system does not allow downloading the tracks, but lucky users can click on the blog post link and find a MP3 file ready to be downloaded for free. What Hype Machine suggests to its users, instead, is direct links to purchase tracks from iTunes and Amazon (with a revenue of 5% of sales made through these click-throughs). Currently, Hype Machine indexes 824 handpicked blogs; the system does not only stream the music, but also repost the textual content. As of 2014, the main features of Hype Machine are main page which aggregates recently posted songs from the indexed blogs, love feature where users add songs to their favorite list (and sub-playlists), popular list where the most ‘loved’ songs of the last three days are showcased along with last week’s populars, zeitgeist to see the best artist, albums and tracks of the previous year, listening history through which users can see the tracks their friends’ have recently listened to, latest where the trends in the blogosphere are tracked, Twitter featuring an interactive music chart of songs posted on the social platform, spy featuring a voyeuristic journey through what other people are listening to, album premieres reserved for unreleased albums, HMTV for most blogged music videos, music blog directory revealing the complete list of indexed blogs (filtered by genre & geography), labs to feature the new products / extensions / features and feed, a customizable view of the blogs, artists, tracks and users the users’ follow. A 2009-survey conducted on 2000+ Hype Machine users reveals that 51% of Hype Machine users are 18 to 24 years old, and 46% of them spend money on MP3s. 61% of the audience attend 1-2 concerts every month, whereas one in every six users on Hype Machine are self-proclaimed music bloggers, DJs and musicians. (Hype Machine 2009 User Survey Results 1) Hype Machine mobile app was launched in 2011 on iTunes Store, with a price tag of $3.99.

Apart from this payment, Hype Machine subscription is free to all users, and the system gains traction by pulling in users looking for new music in the indie scene. This free model, resulting in the prevalence of an audience-oriented structure, not only allows musicians to gain support and build a fan base, but also provides a great tool for professionals in the fields of A&R and publishing alike as many of the tracks found on Hype Machine belong to new artists or songwriters who may be in search of a publishing deal. Hype Machine is an important platform for artists, because it is a chance for those who may not be able to break through the overcrowded music marketplace -yet. The main revenue streams of Hype Machine is click throughs (by purchases from iTunes and Amazon), iOS application purchases and advertising sales. Hype Machine merchandises are also on sale, but the profit coming from these are extremely limited. Why is Hype Machine so significant then? It is the first platform to aggregate not only the music but also the reviews and data of multiple blogs, which eventually gives users a more accurate idea of what’s trending in the blogosphere. The Guardian’s Rob Fitzpatrick recently said that there are currently 1264 music genres, yet targeting a growing market of listeners with a more diverse musical knowledge, Hype Machine is not the most varied platform in genre-wise: (1) In Hype Machine, indie is of course a trend, it is an inspiration and the backbone of the Hype Machine culture; but it is not a rule. The structure offers new music beyond commercial markets, and has relatively become DJ and producer centric for the past few years -the Remix Tracker feature, where users can track and subscribe to all remixes of a particular song available on the blogosphere, launched in 2012 is a signifier
of this. In short: If a user is looking for a lesser-known metal song from 2000s, Hype Machine will not provide any search results; but the same user can be the first one to discover the next big thing in electronica scene by wandering through recent posts.

Jenkins defines seven principles of a new media experience, and evaluating Hype Machine with the guidance of few is crucially important to understand the buzz created around the website. First comparison to examine is spreadability vs. drillability; focusing on the engaging characteristics of the circulation of media content, mainly through social networking sites (SNSs) as well as cultural reflections. Hype Machine is as spreadable as an epidemic, especially among an audience who somehow identify themselves close to the industry. The website has a growing market of listeners with a more diverse musical knowledge, and they are more than willing to share this through integrated SNS buttons with their friends on Facebook & Twitter. Drillability, on the other hand, is a rather more complex metaphor also concerned with devoted practices of engagement. The urge to discover more nourished drillability, as users dig deeper to reach undiscovered parts of a narrative: "Drillable media typically engage for fewer people, but occupy more of their time & energies in a vertical descent into a text's complexities", while spreadability tends to favour shorter-term engagements with a greater audience. (Jenkins 1) The overall narrative of Hype Machine can be considered as spreadable, thinking of how integrated it is not only with most popular SNSSs but also with various other websites (even music streaming platforms) with functional plug-ins, including the Spotify app and Last.fm's scrobbler patch. Hype Hotel, Hype Machine's SXSW-exclusive event taking place for the past three years in Austin, is also an important branch of the brand, offering an offline experience in its Webified culture. Yet this does not mean that only horizontal ripples are encouraged, as some drillable media characteristics are also visible -subtly less, but they do exist. After all, an ideal scenario for a transmedia experience is not about making a choice in between, but to make sure that there are spreadable attractions and drillable complexity: Hype Machine offers its users various customizable music streaming experiences, varying from a casual listening session without few or no interaction to using Lab's experimental apps such as Global Spy (a feature to see global streaming trends on Hype Machine) or Blog Recommender (a tool to suggest music blogs compatible with your recent listening patterns) both of which have drillable media qualities because they are more specific in a sense that they attract and engage with fewer people, and chances are high that this will result in a long-term engagement. Regardless from the duration though, the term engagement is quite crucial to understand the value of a transmedia experience. Such a statement instantly connotates two words: Participation and interactivity, two important elements of a transmedia experience; although Jenkins relentlessly argues that the latter is not necessarily indispensable: "Interactivity has to do with the properties of the technology and participation has to do with the properties of the culture. Obviously, in practice, both may come into play in the same text" asserts when he's being asked to clarify the difference. (1) Also, however self-explanatory it may sound, there are even more blurred lines while discussing participation and participatory culture. In order to clarify the distinctions, Jenkins includes these definitions in a course syllabus:

Fandom refers to the social structures and cultural practices created by the most passionately engaged consumers of mass media properties; participatory culture refers more broadly to any kind of cultural production which starts at the grassroots level and which is open to broad participation; and Web 2.0 is a business model that sustains many web-based projects that rely on principles such as user-creation and moderation, social networking, and crowdsourcing. (Jenkins 2)
The term fandom is no stranger to music industry, thanks to *passionately engaged consumers* of musicians and/or bands. Thereby, it is hard to imagine Hype Machine as a brand that can create that much loyalty and passion; but it somehow does. It is a citizen of Web 2.0, thus supporting user moderation and social networking. Mentioning the existence of participatory culture is also possible: *Hype Machine Union* is a 2012-project where users can apply for an official position (with a little to no payment) at Hype Machine as an ambassador to share the website with their college campus and city. This idea of an offline invasion is quite remarkable for an online platform like Hype Machine, especially when the fact that the whole plan relies on the loyalty of users is taken into account. Fandom, however, is much more intense and the goals of any triggered action is blurry because the urge is what matters most. A great example to illustrate how fandom reinvents itself in the Webified world may be as simple as a tattoo. In a 2011 blog post, Hype Machine team revealed a photo sent from a user called Brett, featuring the phonograph and speech bubble (which happens to be the logo of Hype Machine) tattooed on his arm. Therefore, along with continuity vs. multiplicity principle where Jenkins mentions the elaborate continuities revolving around transmedia narratives, spreadibility vs. drillability pave a way to fan fiction. Developed by Dirty Water Labs studios, *UberHype* is an application born out of the burgeoning culture of Hype Machine. It is a social music discovery tool for android platforms, where Hype Machine has not released an official application yet. Tracking and aggregating the content on Hype Machine, *UberHype* is an unofficial satellite of Hype Machine universe, expanding the transmedia experience to one media further as well as illustrating an example for (rather unconventional) fan productivity.

Immersion vs. extractability reveals another remarkable statement about a transmedia experience, as Jenkins also argues; the is immersed into a fictional world, but a fan can extract parts of a story in various spaces of everyday life. (Jenkins 1) The virtual world of Hype Machine can be concluded as somehow fictional, depending on the definition of the term in a broader sense. Yet what needs to be recognized is the micro world and culture it has created among the music industry, resulting in Hype Machine popular lists becoming a standard for success in indie-influenced genres. To illustrate an example, let’s take a look at a press release from *Big Hassle*, a NYC music PR company representing a wide range of names from Backstreet Boys and Against Me! to Mumford & Sons and alt-J. With an attempt to promote English indie rock outfit alt-J’s upcoming album, Big Hassle serviced an e-press release in July 2014. The release focused on the success of the first single of the album; "Hunger Of The Pine", and one of the most important indicators of this success was justified as following: "This single comes to fans on the heels of the release of the first album track 'Hunger Of The Pine' which reached #1 on the Billboard+Twitter RealTime Chart on June 19, was #1 on Hype Machine". An elaborate look at this release infers that Hype Machine has become an industry-standard as of now, thus it can also be said that the once-traditional industry is now (partially) immersed in the online reality created by the Hype Machine. Analyzing extractability, on the other hand, has rather physical echoes in the Hype Machine universe; caused eventually by the wide use of applications and plugins allowing users to stay connected with Hype Machine in all day & every day, which means that users can extract some parts of the Hype Machine universe depending on their needs. Seriality, in this case, is another principle that is related with this dispersed medium structure: "A serial creates meaningful and compelling story chunks and then disperses the full story across multiple installments". (Jenkins 1) The complete transmedia experience Hypem offers has actually started with the first website launched in 2005; a rather-simple design and aggregating nearly 60 music blogs. However,
the story has been spread to various platforms not by replicating the original but offering medium-specific features adding value to the overall narrative. Hype Machine's iOS app can illustrate a perfect example to discuss the impact of seriality and extractability. With an extremely iOS friendly and functional interface, the app offers a mobile music library shaped by a particular Hype Machine taste and most importantly, an offline mode where the cached favorite tracks can be listened even when the user is not connected to the Internet. The application, like the website itself, allows various playlists with individual micro narratives in a sense; you can click on a genre and listen to a customized broadcast, or stream a particular blog’s latest posts. This definitely makes it easier to find and get music; as the whole playlist of the blogosphere is neatly split into genres and music is searchable with multiple filters (recent uploads, loves, remixes etc.)

Other serial extensions of Hype Machine exist in both online and offline platforms, which is a solid proof of how dispersed Hype Machine is. The online extensions include the Spotify app where users can listen to their Hype Machine feed inside Spotify without the need of a separate browser, and Sonos app where the users can connect their Hype Machine account with Sonos-branded wireless music systems. A now-old fashioned example is Hype Machine’s Last.fm scrobbler, which integrates your Hype Machine streams to Last.fm profile. As it might be recognized, all these tools contribute to different parts of the overall narrative; while an user can read the blog posts on a track in their mobile device, Sonos is more prone to deliver audio streams thorough a hardware. The serial extensions in offline platform vary from events, including Hype on the Go: Handpicked Series where Hype Machine’s favorite DJs, artists and bloggers tour remarkable US and UK festivals with a traveling Hype Machine stage, and The Hype Hotel, the annual event series to host 30k guests in 2014. Aforementioned Hype Machine Union programme is also a serial extension, developing a rather different storyline compared to the others. The worldbuilding principle is linked to principles of immersion and extractability too, as what results in a coherent whole is nothing but particular characteristics that the Hype Machine universe has. How worldbuilding principle is interpreted in this particular case might be different than what Jenkins highlights - the maps and the rules are more applicable to an ARG experience compared to a website streaming new music.

Subjectivity, the sixth principle of Jenkins’ transmedia storytelling, deals with the characteristics of transmedia extensions. After mentioning the power of transmedia extensions to tease unexplored dimensions of the story as well as lengthening the timeline, Jenkins elaborates the subjectivity principle with another statement closely linked to the epistolary novel tradition; where early novelists supported their main storyline with letters and journals, with none or "little acknowledgement of their fictional status". (3) In this case, however, a direct example is impossible to illustrate, yet this secondary character, performing as a supplementary sub-narrative can be interpreted as various ways. Zeitgeist, as briefly explained before, is a rather interesting feature when it comes to evaluating subjectivity, as what it does is to shift the timeline of a platform like Hype Machine backwards where being in-the-now is usually glorified more than anything. Seeing the best music of the previous years is quite crucial, because it is a statement on how fast the content is consumed in online environments and how the worthy ones can actually stay to be revisited. Visually, Zeitgeist functions in a similar way that a micro-site does. Just like Fast Forward, a 2012-project featuring a micro-site which offers new experiences and perspectives to grow in the larger habitat of Hype Machine with the motto "fly through new music". The project is not a success story, but such idea of an immersing and fast way to explore new music is excessively important:
From the latest, popular, or genre channels of blogged music, we show you screenshots of related blog posts, while playing short samples of the songs being discussed. Want to read the whole post? Click the big image on your screen. Like the song? Click the heart to add it to your Hype Machine favorites. Use your arrow keys, space bar, or the > button to skip to the next sample. (From the lab: Fast Forward 1)

Celebrating the launch of the project in 2012, Hype Machine team designed a contest, that adds a pleasantly strange branch to the transmedia experience: Users were asked to download and decorate any of the buildings on Fast Forward’s main page, which can be painted, drawn, printed and scanned back.

The last principle of transmedia storytelling is performance, which is rather self-explanatory except for the fact that those who perform are among the audience instead. In order to draw a community of like-minded individuals (and encourage them to perform), transmedia stories need cultural attractors; such as Hype Machine’s indie-favoring genre structure, or visual aesthetics that can easily be connonated with the Tumblr-y facet of Web 2.0. It can even go to the extremes and become involved with a political or social statement, similar to what Hype Machine experienced during the anti-campaign of Stop Online Piracy Act (SOPA) few years ago, where the team and a group of users protested the "abusive legislation" under the alias Hype Activists. Hence, a performatve aspect was added to the overall transmedia narrative of Hype Machine: the users protested for the future of a platform they feel themselves belong to.

3. Conclusion

Music Industry, in 21st Century, has been building new business models, favoring new platforms and finding more profitable means of revenue; the process of innovation is visible, yet it is far too linear. Breaking the boundaries of this linearity, Hype Machine manufactures a true transmedia experience - a story that is spreadable, dispersed, multiple, serial and subjective. This transmedia-favoring structure is the essence of Hype Machine being "the future of all media". Also, this structure helps to provide new revenue models not only for Hype Machine, but also for the industry itself. Hype Machine has now transformed into a whole universe, with a soul, language and culture, and with its own traditions: Artists who want their music to be heard in the website are able to make genre and style-related assumptions, users are more knowledgeable about how to find their way to maximize their experience in these broad lands.

It is important to keep in mind that however dispersed Hype Machine is through different media outlets and environments in both online and offline platforms, it is still a user-related choice to decide on which parts he/she wants to take or leave, and how to construct a tailor made experience with these. The fact that there is no enforcement in the structure and practice of Hype Machine is the reason why it offers such a powerful transmedia experience. Nothing is linear; the users have options, so does the music industry; and altogether a fluid and dynamic story is being built with a never ending pace. And this comes by no surprise; after all, Hype Machine by essence is a product of transmedia as a music blog aggregator. Thus evolved audience-oriented structure is also vital, as this is one of the most important characteristics that sets Hype Machine apart from highly-commercialized contemporaries like Spotify and Deezer: People are listening
music through them, but they fail to create a culture and overall narrative where the users can live in and look for more in the rather complex consensus of the age of collective intelligence. And the best part is, Hype Machine will never become a monopoly by leading to their extinction. The structure of Hype Machine is nourished with the Webified idea of plurality thus the representational remediation; you can still integrate various listening habits to the Hype Machine platform, or vice versa. The number of mediums/platforms in everyday usage may become less, but all in all, the essence of what makes them unique will survive in an appreciable extent; attributing a representational quality to the act of remediation.

This study reappropriates ideas on transmedia to the infertile territories of music industry / technology studies; in fact Henry Jenkins has probably never have thought that his seven principles of transmedia storytelling can be applied to analyze Hype Machine. However, how powerful Hype Machine embodies these principles is not a coincidence; instead, it somehow serves as a road map, or formula, to shed light on the success of what 19-year-old Anthony Volodkin started in 2005 as a college student who could not find anything new to listen. You can think of Hype Machine as radio or television; your favorite podcast or the genre-wise inadequate version of Spotify, a music magazine or an event promoter; in fact, Hype Machine is all these and more, only slightly better.

REFERENCES


Messiaen, Composer of ‘Awe’: The Proto-Spectralist Pursuit of Divine Fear in Apparition de l’Église éternelle

MIRIAM PIILONEN
NORTHWESTERN UNIVERSITY’S BIENEN SCHOOL OF MUSIC
miriampiilonen2018@u.northwestern.edu

Abstract

Olivier Messiaen’s legacy as proto-spectralist composer has gained momentum since his students Gérard Grisey and Tristan Murail elaborated the genre in 1970s Paris. Grisey and Murail’s practices of graphic representation and digital manipulation of the timbral spectrum were influenced by Messiaen’s tonal innovations and his fascination with the connection between sight and sound. Though Messiaen himself never used the term spectralism, his music anticipates the spectralist movement by sharing its transmutative aspiration. This paper demonstrates that ‘technology’ of sensory-transmutation was present in the early works of Messiaen, citing his 1932 organ work Apparition de l’Église éternelle as a pivotal demonstration of early spectralism. While Messiaen is often considered a composer of mathematical music, his works are frequently sensual and spiritual, shaped by his synesthesia, unconventional theories of time, and deep commitment to orthodox Catholicism. In Apparition a constellation of imagined visual features, emotional sublimity, and suspension of Earthly time aspire toward an encounter with divine awe or fear of God. However, many listeners do not respond to Apparition with its desired level of fear.

In the grandiose timbre of Messiaen’s Sainte-Trinité pipe organ, Apparition aims to achieve three spectral feats: 1) suspension of clock-time, 2) crossing between eye and ear through synesthesia, and 3) transformation of the church from visible vastness to aural vastness. This paper approaches the cross-sensory listening experience of Apparition using score analysis, Christopher Hasty’s projective analysis, and an exploration of imagined visual features. Collectively these analyses examine Apparition’s homogenous texture, absence of melody, trés lent tempo, and synesthetic use of Surrealist simultaneous contrast (the perceptual effect of side-by-side contrasting elements) as Messiaen’s means to achieve divine fear. This paper concludes with a discussion of the pluralistic analytical interpretations that may arise.

Keywords: Messiaen, spectralism, music analysis, music perception, theology
Introduction

Olivier Messiaen’s mother was the poet Cécile Sauvage, a woman of deep depression and mortal dread (Schloesser, Visions). While pregnant with Olivier, she imagined his birth stalked by death:

*Child, pale embryo, asleep in your water bed*
*Like a little god in a glass coffin, lying dead…*

The congruence of birth and death resonated strongly with the young Olivier; it would become a metaphysical obsession and compositional motif (Dingle & Simeone).

Mother and son each contemplated the meaning of Earthly existence, which seemed cruel and ephemeral. Olivier found meaning through religion, devoting himself to Catholicism from an early age. Cécile, however, was agnostic. Her life – embroiled in war, deaths of children and friends, and the metaphysical questions that plagued her – ended after a period of illness and depression when Olivier was 19 (Schloesser, *Visions*). Near the end of his life, Messiaen recounted, “There has been only one influence in my life, and it happened before I was born… I believe it was that which influenced my entire destiny” (Dingle 2). He spoke not of music, not of God, not of the countless artistic and philosophical paragons he admired, but of his mother, Cécile Sauvage. Messiaen’s 1932 organ composition *Apparition de l’Église éternelle* has been suggested as a grief work written in her honor (Schloesser, *Visions*). *Apparition* also demonstrates early and transparent experiments with sensory transmutations: the artistic means by which Messiaen would perform his and his mother’s creative pursuit of eternal life. Apparition, as I will show, is a composition meant to evoke mortal dread and capitulation to a Catholic God. Messiaen inherited Cécile’s artistic preoccupation with life and death: he possessed a "melancholic yearning for immorality" (Schloesser 152) which he pursued through creative means. Inspired by his readings of Ernest Hello’s theology, he aimed to cross from simple fear (*la peur*) to divine fear (*la crainte*) as a means of self-transcendence. Believing that a musical event could evoke that crossing, he aspired toward an aesthetic experience that exceeded the capacity of ordinary music listening (Schloesser, "Fear"). Pushing beyond the score, beyond the sound, and beyond the Earthly means of music, he wielded synesthetic color-imaginings, space-as-sound, and suspension of clock-time. Messiaen intended for these sensory transmutations to precipitate a particular level of fear, that of divine fear or fear of God (Schloesser, *Visions*). Decades later, Messiaen’s students, the spectralists, would employ the same sensory-transmutative practices toward a non-religious aesthetic goal with their graphic representation and digital manipulation of the tonal spectrum.

This paper aims to unite the compositional worlds of Messiaen and his spectralist students through their shared cross-sensory creative practices. Primarily, I will elucidate Messiaen’s intended sensory transmutations (‘spectral feats’ as I will call them) through the organ composition *Apparition*: 1) suspension of clock-time, 2) crossing between eye and ear through synesthesia, and 3) transformation of the church from visible vastness to aural vastness. In the process of that elucidation, it will become clear that Messiaen’s intended transcendental effect may be lost on the way to listening ears (particularly those who, like Cécile, do not believe in a Catholic God). Secondarily, it reveals that the spectralists, though not explicitly tied to religious or secular sublimity, pursue the same subjective sensory transmutations. I will conclude with a discussion of the role of the analyst in elucidating spectralist cross-sensory experience in light of Messiaen’s pursuit of divine fear.
**Spectralism Undefined**

Spectralism (or spectral music) originated in Paris at the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) by the Ensemble l’Itinéraire in the 1970s. The term was coined by Huguées Dufourt and reinterpreted more broadly by Messiaen’s students Gérard Grisey and Tristan Murail (Anderson). So, to begin, what is spectral music? What techniques are involved in its composition? What distinguishes its sound?

An immersion in the spectral corner of the music theoretical literature reveals a condemnatory concern: there is no working definition of spectralism. In Julian Anderson’s Provisional History of Spectral Music he writes:

> There is no real school of spectral composers... As far as I can ascertain, the nomenclature 'spectral' is regarded by virtually every major practitioner of the trend as inappropriate, misleadingly simplistic and extremely reductive. The use of spectra, whether harmonic or non-harmonic, is only the most superficial feature of the music of these composers. A much more fundamental concern... is with the conscious composition of the degree of predictability or unpredictability in their music, and a consequent fascination with the psychology of perception. Finally, most composers involved with this tendency shared an inclination to transcend the limits of parametric composition: the serial obsession with devising separate or related charts for pitch, duration, intensity, dynamics and timbre is replaced with a fondness for attempting to abolish the distinctions between these phenomena (8).

Though there appears to be a misunderstood or indeterminate 'spectral' dictum, Anderson’s introduction suggests the elision of Messiaen and the spectralists in several key ways. First, Anderson de-emphasizes compositional usage of spectra but is adamant about the practice of abolishing distinctions between musical phenomena. I hypothesize that the use of musical spectra, whether harmonic, timbral, or otherwise, also underlies Messiaen’s desire to traverse between musical parameters. This will be explored through score study.

Second, Anderson’s emphasis on listener psychology summons Messiaen’s well-known interest in music perception (Samuel; Schloesser, Visions). This paper will explore listener 'throwing-forth' of expectation using Christopher Hasty’s analytical tool projective analysis. Here I will explore Messiaen's first intended spectral feat: the suspension of clock-time.

Finally, the notion of "non-harmonic" spectra suggests non-musical interpretive possibilities. I will show here that the piece attempts two further spectral feats: 1) crossing between eye and ear through micro- and macro-displays of Surrealist simultaneous contrast, and 2) transformation of the church from visible vastness to aural vastness.

Collectively these explorations will show that Apparition uses 'technology' of sensory transmutation to achieve a proto-spectralist effect, despite predating the computer. Apparition’s loudness, homogenous texture, slow tempo, and lack of melody transform into colors, visions, and bodily discomfort, intended to evoke the piece’s highest spectral aspiration: crossing from simple fear to divine fear, or fear of a deity. Gérard Grisey described spectral music as fundamentally "liminal" (Anderson 8). In my closing remarks I will comment on the analyst’s role in exploring compositional practices that feature both liminal cross-sensory effects and reliance on listener perspective.

*Apparition* appears to lack formal structure; its texture never changes, its motives never vary, it has no cadences or sectional delineations of any kind. The piece’s only
Score Study

Apparition’s core is its timbre. The pipe organ played by Messiaen is a machine of transcendental size and indefatigability, a Cavaillé-Coll grand organ. It spans over six octaves, can produce volume over 85 decibels, and has a distinctive klang or overtone structure. Though the Cavaillé-Coll’s stoppered pipe design eliminates many of the higher harmonics (those that are closest together and produce the strongest dissonances) the others cannot be completely quieted (Shannon). Already the crossing from timbre to harmony is present; Apparition’s chords are colored by the harmonically rich voice of the pipe organ (McAdams & Giordano).

Apparition is nothing but ten minutes of homophonic chords. The score has no time signature but its rhythms have been corralled into bars (Figure 1). The same rhythmic figure repeats in each bar (eighth-quarter-held note) but stretches and shrinks to accommodate the measures’ changing durations. The chords vacillate between dissonance (the “bright colors” of the 2nd, 3rd, and 7th ‘Modes of Limited Transposition’) and “cold and hard” open fourths and fifths. Two dissonant chords in an eighth-quarter note rhythm (the “iamb” figure) are followed by a “double long” open consonant chord (Figure 2) (Messiaen, "Complete Organ Works").
dynamics are an immense crescendo to the midpoint followed by a progressive diminuendo. Climactic whole notes mark the middle of the piece: ascending Ab minor, Bb minor, and C major chords with plenty of Messiaen's parallel fifths, fourths, and octaves (Figure 3). The swell from piano (soft) to fortississimo (very loud) is imperceptibly gradual, as is its return to quiet. Apparition is unconditionally non-melodic, a conspicuous oddity: later Messiaen would announce in his *Technique of My Musical Language*, "Supremacy to melody! The noblest element of music, may melody be the principal aim of our investigations" (Messiaen, *Treatise*, 32).

Furthermore, the *Très lent* tempo exaggerates the piece’s heavy soundscape; listeners do not perceive rhythmic organization when music is performed too slowly. An estimated lower tempo limit is 2.5 seconds between sounds (McAuley), which *Apparition* reaches in every bar. How would music without its noblest element behave? Like spectral composers, Messiaen was highly concerned with the psychology of the listener (Samuel). Denying the listener their melody produces a peculiar effect. This concern will be explored through listener projection.

**Projective Analysis**

In the years after the death of Messiaen’s mother, his conflicted relationship with life’s ephemerality bled into his music. Drawing on the philosophy of Henri Bergson and Alfred North Whitehead, Messiaen developed a distinctive approach to rhythm and meter that rejected discrete, atomized musical moments in favor of the ongoing flux of duration – that is, time as it is lived and felt, not measure by a metronome or clock (Hill & Simeone).

And so, it appears Messiaen’s music is suited to Christopher Hasty’s innovative work with music and temporality, outlined in his book *Meter as Rhythm*. The book asserts that music analysis frequently explains music in terms of what it has become rather than by the process of becoming, excluding temporal phenomena and treating music as a static object. Hasty proposes a new analytical tool, *projective analysis*, that explores the listener projection of expectation, based on the same theories of time that captivated Messiaen: those of Bergson and Whitehead. Hasty emphasizes that the 'text' from which an interpretation is realized must be an acoustic one, that is, a recording or performance (Hasty). This study conducted several analyses with different performances but will only
report here on Messiaen’s own recorded performance (Messiaen, “Complete Organ Works”).

The first four bars introduce the iamb and double long rhythm. The first projection, R – R’, is an eighth note projection, immediately disappointed by a quarter note duration. A second projection, Q – Q’, anticipates the eighth-quarter rhythm, the iamb, again frustrated by a half note duration. The first fulfilled duration is the entire rhythmic unit, the iamb and double long, though the slow tempo suggests that a consistent projection may not occur in the first four bars (Figure 4).

![Figure 4](image)

The iamb presents the first rhythmic confusion in the piece. Depending on the performance style, the number of iterations of the iamb that have already been heard, and the chords themselves, the listener may interpret the beginning of each phrase differently. Figures 5 and 6 illustrate two possibilities. In Figure 5, the eighth note is heard as beginning which suggests the original projective analysis. However, if the eighth note is heard as anacrustic the projection R – R’ will not be heard until the quarter note attack (Figure 6).

It is likely that after several iterations of the iamb, the initial attack as beginning will become the main projection for the remainder of the piece. Still, the slow tempo continues to suppress perception of rhythm and therefore projection of the iamb. It is equally likely that a first listening to this piece will not produce a projection of the iamb at all. Instead, the listener may simply perceive disconnected chordal attacks. This is supported by Paul Festa’s documentary on the piece, in which many listeners appear not to notice the iamb (Apparition of the Eternal Church). A complication for each of these projections is the pedal, which re-articulates a syncopated pattern during each manuel half note. One interpretation of the pedal’s effect is that it will go unnoticed beneath the foreground material. Alternately, a fourth projection would begin with each pedal attack.

In Apparition’s distorted durations – its long long long feeling of now – there is constant tension. Without melody there remain only isolated chordal attacks, the repetitive
persistence of the chords themselves, the looming crescendo and swooning decrescendo, the great contrast between dissonance and consonance, the rhythmic beating of dissonance in absence of a more conventional rhythmicity. And, perhaps, an array of thoughts, feelings, and visions within the listener’s imagination.

In the remaining section, I will conduct an inquiry beyond the score, substantiating the proposed visual features of the piece and demonstrating the proto-spectralist sensory transmutation from aural to visual and aural to spatial.

Visual Exploration

*Apparition* wants to achieve Cécile and Olivier’s artistic goals and yet there is a chasm between the piece’s elevated aspiration and its lowly material means. Though *Apparition* desires an encounter with divine awe, it has only Earthly pitch, harmony, timbre, dynamics, force, and power. Can it produce an awe-ful effect through that which it aims to transcend? Messiaen would say that it does by making the listener cower, drawn from Ernest Hello’s notion of crossing from ‘simple fear’ (la peur) to ‘awe’ (la crainte) as a means of self-transcendence (Schloesser, "Fear"). I will show that the cross-sensory features of *Apparition* were designed to produce simple fear, from which the listener may ascend to divine fear.

First, space becomes sound. The organ is built into the walls of La Trinité church. As the pipes bellow, sound waves fill the chapel – the resonance chamber. As Messiaen plays the body of the church, visual vastness transforms into auditory vastness. The pipe organ’s volume charges the chapel, rendering it musically and spatially ‘full’.

Second, sound becomes vision. Messiaen’s synesthesia was viewed, at the time, as a rare pathology and allegory for achieving mind-soul unity through multisensory traversal
(Schloesser, *Visions*). His imagined colors facilitated a musical semblance of simultaneous contrast (Chevreul). For Messiaen, *Apparition’s* dissonant chords (his 2nd, 3rd, and 7th modes) inspired bright colors, while consonant fifths inspired cool ones (Messiaen, “Complete Organ Works”). Andrew Shenton writes, “For Messiaen, natural resonance (the harmonic series) was the aural equivalent of simultaneous contrast, and he believed that his musical colors were perceived differently depending on what surrounded them” (51). This aligns with Messiaen’s love of Surrealist and Cubist art, which make use of the same color-theory phenomenon (Shloesser, *Visions*).

The color associations Messiaen described in writing and interviews were vivid and precise. For example, the second transposition of mode 3 inspired horizontal layered stripes: from bottom to top, dark gray, mauve, light gray, and white with mauve and pale yellow highlights – with flaming gold letters, of an unknown script, and a quantity of little red or blue arcs that are very thin, very fine, hardly visible. Dominant are gray and mauve (Samuel 64).

These complex color combinations – the 'bright' imaginings of Messiaen’s dissonant chords – are contrasted with the cold colors of open fifths and fourths, producing *Apparition’s* contradistinctive palette. An example of how this would appear in the score is depicted, by vulgar necessity, in red and blue (Figure 7).

![Figure 7](image-url)

While the opposition of warm and cold colors produces a moment-to-moment simultaneous contrast, a larger, macro-display of simultaneous contrast is also present: the opposition of the Divine Church’s beauty and violence. In the piece’s accompanying poem, Messiaen wrote of the Book of Revelations’ gems of twelve colors, the twelve chromatic tones of music, and the martyrdom of human souls as stones in the Eternal Church’s foundation, all in a single motif: the *living stones*.

*Made of living stones,*  
*Made of stones of heaven,*  
*It appears in heaven:*  
*It is the Spouse of the Lamb!*  
*It is the Church of heaven*  
*Made of the stones of heaven*  
*That are the souls of the elect.*  
*They are in God and God in them*  
*For the eternity in heaven!* (Schloesser 171)
Apparition’s beauty and terror swarms out of the vast resonating chamber of the chapel and blast the listener with its unrelenting din. Just as the organ timbre burgeons the notated harmonies into lustrous cacophonies, the harmonies want to inspire Messiaen’s imagined colors. Just as the organ renders the space of the church present, active, and full, so in the crossing of senses Apparition wants to beckon forth divine fear. Its title wants to reveal all: the piece is an Apparition of the Eternal Church, staggeringly beautiful, made of martyred human souls. Messiaen wrote it for performance during Catholic Mass to terrify the congregation into capitulation. Apparition is intended as an elegy, sensory coalescence, and divine sonic boom.

But Cécile Sauvage, the listener who Messiaen may have most desperately wanted to achieve the full traversal, did not believe in God.

Conclusion

If it remains unclear, let me reiterate: Apparition can be unpleasant. It is voluminous, homogenous, and unrelenting. It resembles the sound weapons used to disperse rioters and frighten civilians during wartime (Goodman). It is heard in the ears, seen in the mind’s eye, and felt in the body. At its highest level of aspiration it terrorizes to the God-fearing core. Those who subscribed to the Catholic God’s command of sacrifice and martyrdom may have achieved the crossing from simple fear to divine fear. And yet Messiaen’s Catholic God would have played no part in his mother’s experience of Apparition.

As overwhelming and exhausting as the piece is, a listener may never feel the feeling that it demands of them, and seems to offer in itself. Does this mean their listening is a failure? With such a clear intention set forth in the title, what does it mean if the listener fails to achieve some fetishized out-of-body sublimity, surpass bodily threat, imagine beyond magnitudinous extremes, or dip out of the durational demands of Earthly conceptions of time? Though Apparition tries to force the listener elsewhere, toward awe, its ability to do so appears to depend on the philosophical privilege granted to it by that listener. Does it require belief in God? To pivot away from Catholicism, does it require synesthesia? Does it require a live performance in Messiaen’s church? Without these, can Apparition do anything more than blow air in the ear canals, thud in the chest, and leave one writing tediously speculative analyses?

And what is my, the analyst’s, role in the piece’s interpretation? Should I have to alter myself to appease that ghastly Messiaen? Which of us tunes to the other? Similarly, would an analyst (or listener) without awareness of Gérard Grisey’s use of the spectrogram hear the illusory trombone of his orchestral piece Partiels? Does Tristan Murail’s polyphony ‘sound like’ the process of FM, RM, and AM synthesis (Fineberg)?

And so, I must insist: Apparition as a piece of music is just that, a piece of music. As much as Apparition wants to push its listener beyond the senses, it can only transmit sensation. Its sensory transmutative aspirations – those that inspired the spectral composers – are equally at the mercy of interpretation. Perhaps then, the critical statement spectralism makes is that the musical and non-musical spectra are indeed conceptually isolated worlds – expeditions apart – but are interconnected by composer, performer, and listener perspective. The role of the analyst is therefore not elucidatory but revelatory, shining light into a corner of possible interpretations while acknowledging the multiplicity built into musical experience.
REFERENCES


Establishing a Signal Processing Methodology for Multi-track Mixing Education

C. BARKIN ENGİN
BAHÇEŞEHİR ÜNİVERSİTESİ SES TEKNOLOJİLERİ YÜKSEK LİSANS PROGRAMI
barkinengin@gmail.com

Abstract

The fundamentals of multi-track music mixing - the first phase of traditional audio post-production - are fairly apparent on the surface, yet matching the aesthetical ideas with technical knowledge require a certain amount of trial and error period and an advanced critical listening approach along with an efficient organizational logic. From the educational perspective, starting with the basic session management procedures, proceeding to genre appreciation and introducing related signal processing tools and techniques are universal approaches to the audio post-production curriculums. Nevertheless, theoretically infinite signal processing capabilities of a modern day digital audio facility entails potential pitfalls for students, over-processing being the most common and destructive one. The common non technology dependent problems such as perspective loss and ear fatigue are still relevant and improper spectrum and dynamic management supports both of the situations.

Over-processing dramatically affects the reception of any musical composition. Consequently, introducing an educational methodology for signal processing might be helpful in order to avoid cumulative dynamic, spectral and spatial damage done to the original recordings, which are often unsolvable in later stages of mixing or in mastering and to minimize the risk of aural perception based misjudgments.

Introduction

Like its counterparts in other post production disciplines, multi-track mixing is a highly aesthetically driven and creative process with certain technical requirements. There are not only genre related approaches or commercial concerns regarding the micro and macro structure of a mix. One can also add her/his perspective to the equation and develop a unique style, which may become the sonic characteristic of a musical genre, or may trigger the establishment of a completely new one. Therefore any educational methodology to musical post-production should not be considered as absolute sets of rules. They must be understood as an attempt to explore all stages of a mix with particular working order and assigned limitations in order to reduce previously mentioned risks and to offer a better picture of the progress. Once the student becomes familiar with all kinds of decision making flow, the methodology can be modified or completely abandoned.

Although the foundations of the multi-track mixing in film and video production is analogous, present layers, their functions and interrelations are completely different than music mixing. The core idea behind the methodology might be translatable to audio post
for moving images, yet the term over-processing has to be covered in an entirely new context and the parameters of the related signal processing tools have to be adjusted accordingly.

The proposed methodology aims to divide the overall processing into three different conceptual stages; initial, contextual and final / creative processing. Certain limitations, goals and problem solving methods will be specified for each stage. Dynamic and spectral manipulations have priority in the initial stage and other forms of audio manipulation will be discussed later in the contextual and final stages. This processing standard is compatible with the traditional aspects of multi-track mixing and preserves the essentials of the recent global education strategies. Additionally, lesser-known processing techniques, such as parallel compression and feathered equalization, are also given significance due to their more transparent sonic outcomes, although transparency is not a desired feature for every musical style.

**Initial Processing**

The initial stage of the mixing process covers respectively the organizational procedures, session management, basic balance / panorama adjustments and early spectral and dynamic signal processing. Becoming familiar with the overall composition and the individual tracks, assuming that the recording is not done by the mixer, is extremely important. The very first reaction to the composition dramatically defines the musical goal and enables the mixer to organize an efficient workspace. Session management starts with necessary software adjustments, labeling of the individual tracks and grouping the related tracks in sequential order. A logically organized session not only saves time during the later stages and revisions, but it is useful to develop a habit of making sense globally. Recent multi-track software presents several features to ensure an easier and safer process. Inputs, outputs and busses must be enumerated (or labeled) for rational routing capabilities. Buffer size must be set to higher values than recording to increase the computing power required for processing large numbers of tracks (Izhaki 160). Delay compensation option must be activated to keep the synchronization between tracks stable regardless of the latency caused by plug-ins or hardware inserts. Post-fader metering is also crucial for a distortion free gain scaling with enough head room to prevent potential inter-sample clipping (Mason). Panning laws must be set according the desired reproduction method, stereo or surround (Izhaki 188).

In cases, where a single sound source is captured by several close or distant microphones, and/ or recorded directly using DI boxes, this is the first chance to determine the primary and secondary tracks and establish a balance hierarchy between them. The question of hierarchy is a question of the desired spectral structure in most of the cases. It is important to keep in mind that different microphones, microphone positions, preamps or any other variation in the recording chain would result with a slightly or dramatically different frequency spectrum. Using two or more sources may result with a fuller or more complete spectrum, yet it is open to experimentation and also can be judged according to the function of the specific instrument or instrumental line in the composition. Associated tracks can be routed to a mono or stereo sub track dependent on the panoramic preference. Using sub tracks offers the possibility to manipulate the combined signal in addition to individual control. A sub track is also a useful device for the final adjustments at the final/ creative stage of mixing. Additionally, the mixer may eliminate some of the tracks, which might be poorly recorded or not relevant to the context.
Using the spectral variables of the same sound source introduces the concept of phase, which is the time relationship between waveforms measured in degrees (Izhaki 163). There are two types of phase relationship; instruments with single sound source and multi sound source (such as drum sets). In both situations, manual or automatic phase alignment might be necessary to avoid the side effects of comb filtering; however occasional comb filtering could be used as an effective tool to achieve a significant type of texture. Another waveform related action during this stage is checking edit points and applying / adjusting crossfades, trimming the unnecessary portions of the audio and handling the attack and decay parts with fade-in and outs. Global fade in and out should be performed at the mastering stage. Panning, placing the individual tracks in the stereo / surround field can be realized simultaneously with rough level settings, since there is a perceptual correlation between them. An instrument with its own space will be masked less by the other elements, though it is important to be aware of the fact that panning strategies are not sufficient enough to entirely fix the problems caused by poor orchestration. There are several conventions about panning of the certain instruments and human voices, especially in stereo systems, yet there are many examples which deviate from the norm. At this stage, no automations will be applied to both parameters.

The closing section entails basic spectral and dynamic management. Equalization at this point only functions as a corrective device, rather than shaping the sound to fit or feature in the mix. The target frequencies are the unwanted resonances and the frequency bands which are not in the range of the specific musical object, so it is a cutting specific operation. The bandwidth of the resonances may vary; therefore having an equalizer with flexible Q options is considerable. One should avoid extreme cuts and limit the range between -1 and -3 dB, unless it is a very problematic situation. Feathered equalization may produce safer results, since it removes also the harmonics of the resonance band, but generally the equalization is applied in lesser amounts (Owinski 42). On the other hand, introducing less artifacts and coloration is essential. All of these factors are decisive for choosing the right type of equalizer. Due to their operating principles and parameter wise limitations, analog equalizers are less likely to be used at this stage.

The same principle might apply to the selection of compressors. Using a compressor is generally a genre and/or performance influenced decision. If it is aesthetically or technically beneficial to use a compressor, selecting a transparent one with moderate attack and release times, dictated by the envelope of the instrument / sound, and modest ratio threshold relationships causing no extreme gain reduction would be preferred. The order of spectral or dynamic processor to be the first insert in the chain is a source and perspective dependent case. A compressor might follow an equalizer, if the track contains redundant frequencies. An obvious example is human voice; otherwise the compressor would be busy reacting to unwanted frequencies such as the decrease in the bass response caused by proximity effect (Eargle 17). A contrary example could be the complete drum set, whose range covers the whole human hearing; therefore the choice would be a matter of goal. If necessary, gating, sample triggering and side chain compression should be applied at this stage. After the processing is complete for every track, a revision to pan and balance might be required.
Contextual Processing

Contextual processing, the second stage of mixing, should be considered where advanced processing may have take place, a fact that often shaped by the aesthetic requirements of the genre. Micro level frequency spectrum can be modified further in order to achieve a specific macro texture. In other words, fit and feature oriented radical equalization may have take place. It is impossible to generalize certain recommended settings. On certain musical examples the genre is associated with extreme timbre manipulation, and for some of the cases the opposite, purist approach must be taken. Using parallel distortion might, where the unprocessed and distorted version of the same signal is mixed, produce interesting textures or can accomplish the featuring action more efficiently (Senior 194). Similar rules apply to the question of further compression or limiting. The dynamic strategy is again aesthetic preference dependent, but the technique called parallel compression enables the mixer to acquire bolder compression with less harm to transients and stereo image, and avoids pumping and breathing. This is done by mixing the heavily compressed and uncompressed versions of the signal. By changing the ratio of the balance, more or less compression can be obtained (Katz 134). Parallel compression is significantly advantageous on instruments with broad spectra and also helps to underline a certain layer in dense arrangements. An appropriate example is the drum sets used in rock music and its derivatives. It is useful to decide at this point, if employment of de-essing or multiband compression processors is essential before proceeding to the next chapter.

The next crucial step is the introduction of the time based effects. Adding depth and dimension to the tracks not only gives extra "body" to the recordings (particularly important for close microphone applications), but also enhances the dramaturgy of the composition by creating another form of hierarchy, thus changing the reception of the mix. Certain genres use time based effects to achieve ambiguity or apply gestures. Unquestionably, reverb and delay effects are the most frequently used types of the time based effects family. It should be noted that instead of using this type of effects as inserts, creating FX channels using aux channels gives more control on signal flow and provides better system performance due the possibility to route several tracks to the same effect bus. Throughout the history, the choice of the reverb or delay type as well as the percentage of their contribution to mixes is frequently dictated by the genre expectations and audio trends of the time. Regardless of the types, settings and the output level of your time based effect choices, supplementary spectral modification of the effects are an inseparable part of the processing agenda at this moment. The rest of the time based effects; chorus, flanger and phaser, might have variable functions in the mix. From simple intonation enhancers to extreme forms of processing, whatever role do they assume, an equalizer might follow for compensation. In mandatory cases, noise reduction and pitch correction should be applied approximately during this period, since we would have a more established mix with a clearer direction.

Presuming the general processing is done, it is time to recheck the recent gain scaling. It is vital to remember the multi-track software in use might not indicate peak clippings occurred on inserts and due to the internal scaling the fader may not exceed 0 dBFS; meaning no actual visual evidence of clipping is present. Individual meters of plug-in and hardware inserts have to be checked for safety. The next phase is creating automations, commanding changes in time for selected parameters. Often the nature of the sound source, arrangement decisions or performance instabilities demand some form of automation or it may be solely a matter of creativity. Unless it is absolutely mandatory
to create automations during the early stages of the mix, it would be useful to wait until this point. Volume, panning and internal routing parameters are usually subject to automation. One advantage of plug-ins is the automation compatibility of their parameters. In simplest cases, different bands of equalization may be active during different parts of the composition or the thresholds of compressor or gates may be changing according to the dynamic structure of a specific passage and so on. A dynamic and lively mix could be achieved by bringing new perspectives to the concept of parameter automation.

After realizing necessary automation adjustments, observing the global peak and RMS levels at the master bus are highly recommended. The maximum peak level desired by the mastering engineers are variable and requires consulting, yet having global peaks around-3 / -6 dBFS is the most common approach. RMS levels gives information about the dynamic range manipulations of the individual tracks, thus giving us the chance to reconsider our dynamic strategy before entering the final stage of mixing. It should be also reminded that in the majority of the musical mixing projects, a spectral or dynamic processing at the master bus may bring more harm than actual support. Unless you have a strong aesthetical or technical motive, hesitate to process the master bus or at least delay your plans until you test your mix on different playback systems and figure out if your reasons are still there. Nevertheless, a mixing engineer should be closely familiar with the mastering process and must have foresight about what kind of processing should be reserved for that part of musical post-production. Testing the high resolution mix exports on a variety of professional and consumer quality systems enables the mixer to observe whether the essentials of the mix translates sufficiently or not. At the same time, it helps her/him to understand if any anomalies exist due to the poor frequency range and response of speakers, wrong speaker placements or inconvenient room acoustics in the studio. Spectrum analyzers are useful devices but aural perception comes always first. Once the data from the tests are collected, it is time to move on to the final stage.

Final / Creative Processing

The last stage of the proposed mixing methodology is the hardest one to conceptually generalize and it is highly resistant to restrictions. This stage consists of approximately two sections. Inevitably, the first section will cover the revisions, which are noted after the listening tests. This may be limited to slight adjustments in the present signal chain, or may require radical changes or a few additions. Classic examples of addition are specific frequency boosts to a dominant element in the mix to complete the macro textural structure, or the need of extra gain reduction for more stable dynamic behavior. So far, only the fundamental signal processing tools has been covered, yet there are many single tools and combined engines available which can be creatively used in a mix. They may be variations of the essential palette of tools such as dynamic and step filters, enhancers, or they may belong to unique categories such as bit reduction, Doppler shift, pitch shifting and so on. Thus the application of them to the mix is a context specific act. As it is also valid for the previous stages, A/B testing reveals if the additional signal processing adds a value or is the potential consequences not worth it. Some of the processing tools will produce different sonic results regarding their position in the signal chain; therefore a special care must be given to the signal flow experimentation. Some "extreme" processing tools might perform better on a FX aux track, since interaction in the chain may cause undesirable outcomes.
Second section is reserved for creativity and imagination. Once the mixer is done with the construction of advanced micro and macro relationships, it is time to add subtleties and nuances to the composition. Potential musical gestures are infinite; however some common examples (or clichés) can be discussed. Temporal or continuous motion of the elements, achieved by manual or auto panning, adds interest to mix, though not every listener will appreciate it due to the narrow nature of stereo sweet spot. Abrupt or gradual spectral changes in the tracks are increasingly popular especially in popular music genres and commonly used as a transitional function in the arrangements. Dynamic contrasts might enhance the dramatic structure, making sectional volume automations is not a rare tendency. Time based FX automations might add sustain to the decay parts, making a smoother disappearance possible. Human voice is exclusively suitable for diverse range of processing. Automated audio sections may underline the lyrical content. Constant modification leads to intriguing voice textures or may represent unearthly qualities. Macro level processing such as complete filter or time based FX application to an intro / outro etc. can be performed on mastering as well. Although an ordered processing strategy minimize the risk of over-processing, it is not completely eliminated. One has to refrain from quick decisions and always evaluate the mix again with fresh ears and improved perspective. Experience and advanced technical knowledge is required to speed up the decision making process, however healthy aural perception is mandatory for everyone.

The key factor for not getting lost at this creative section depends on several conditions. The individual’s ability to develop an analytical listening approach to musical productions has a definite pivotal role for aesthetical establishment and has to start right away at the early stages of the education. Secondly, continuous questioning of every processing wise decision is a must. Over producing may have similar consequences with over processing or they can merge into a single giant problem. Problematical recordings necessitate bolder actions and bolder actions mean compromises of some sorts. Still the processing strategy must be planned without forgetting the promises of mastering. At the end, mixing should support the composition and performance, not degrade them or make it less comprehensible for the audiences.

Conclusion

By establishing a guide for a logical work flow and sets of restrictions to signal processing, it is expected that a student with no previous experience may explore the inner structure of multi-track mixing in detail, and can get a significant response from every signal processing activities, even from subtle adjustments. The three different stages and their subsections can be modified according to the genre specific requisites, be translated to analog only systems and can be elaborated with specific information for complex instruments / sound sources such as drum sets or orchestral sections.
REFERENCES

You Are a Target Market: Using News Media and Advertising to Create Generative Sound and Visual Art

ALAYNA AINE HUGHES

BERKLEE COLLEGE OF MUSIC, VALENCIA SPAIN

ahuughes@berklee.edu

Introduction

You are a Target Market is a multimedia, multichannel installation that aims to present an artistic representation of the visual and aural barrage that is American advertising and media. The installation was built with programming using Max and generates sounds and visuals on a random basis to present a unique experience for the audience.

Concept

My concept for the sound design of this installation was to create a multichannel sound installation to accompany four walls of video projections. I reviewed numerous commercials from Advertising and news archives and selected videos and sound bites with key phrases and themes to use in my project. I decided to explore for sub-themes for the installation: Media, Gender, Politics, and Consumerism. For the different categories, I would use archive sites to search certain keywords representative of each subtheme. I would begin by brainstorming words that I associated with each category and search those words or phrases. For example, for the Gender portion, I would use keywords such as: sexism, women, women and advertising, and women and politics. For the Consumerism portion, I would focus on searching for words such as: buy, sell, black friday, consume, and capitalism.

Sound Design and Programming Methods

After collecting sounds, I then chopped them up into small samples and phrases. I wanted to create a diverse soundscape, so I chose some of the sound bites to manipulate with effects. The programs Soundhack and SPEAR were excellent for time stretching and harmonic manipulation and yielded some interesting results.

Although I programmed my project entirely in Max, I used programs such as Ableton Live, Sound Hack, and SPEAR to manipulate some of the sound bites that I had collected. For the programming of the sound portion, I built separate patches in Max for each section. Because I wanted my audience to experience a cacophonous effect while viewing the project, it was important to vary my programming in each patch.

I needed to devise several methods within Max to generate my sounds in a unique way. After creating a prototype patch, I liked the idea of using variable volumes on the sounds in order to make different portions audible to the audience over time. To achieve
this, I programmed a simple sfplay object connected to a gain slider. I then made a sub patch to connect to each gain UI object that moved the volume up and down using a line object with varying times in milliseconds for each sound. To vary the frequency that each sound was played back, I put a different number for each metro object that was connected to send a '1' to the sfplay objects of each sound. By using these methods throughout all of my audio patches, I was able to create a continuing, varied soundscape. The end result of this programming was interesting, as it allowed the listener to experience the installation for some time while hearing different bits and pieces of the same sound files as well as some sounds more accentuated than the other.

For other patches, I programmed variations to change the speed and location of other sound bites. For example, on a set of samples in my 'political' section, I used the 'counter' object to speed up or slow down samples. In the 'consumerism' section, I used cue sub patchers to trigger different points of samples as well as slow them down. I used these sounds being manipulated in real-time in Max in conjunction with the designed sounds and unaffected sounds to create an undulating soundscape varying in volume and texture.

Music Composition

In addition to the sound bites that comprised the bulk of the sound design, I also composed music for the installation. I composed the soundtrack using Ableton Live with the intention of creating musical sequences that would be triggered randomly. I composed the music in the genre of electronic/rock/breakcore that I enjoy making and integrated the sound bites as well as manipulated sounds that I had collected. I wanted to make ten pieces around one minute long that all had a different and dis-jointed feel.

To play back the music in a random fashion, I programmed a patch in Max to trigger a bang sent to randomly select a number associated with each music sound file. A piece of music would then play and after the sound file was finished, it was programmed to trigger the bang again to randomly select a piece of music.

Multichannel Programming

To gather all of the sound together in one place where I could control the level to each of eight speakers separately, I built a patch. I created a section for each theme and created a 'receive~' object to send into a gain slider so that I could independently control the volume of each section that I was sending. Out of these gain sliders I sent the signal to my multichannel area where it would be sent to the dac~ into whatever channel that I wished. For example, for the gender sounds, I created 'send' objects in the gender sound patch, which when opened sent the sounds into my multichannel patch and were received and transmitted into my gender volume control. The gender sounds were then send via a send~ gender object into a receive~ gender object, which were then sent through a level of 0.5 and sent to a channel of my dac~ object. I created the same process for the music, but reserved channels one and two for music with the intention of only playing music from a stereo pair of speakers located in the center of the room.
As for the rest of the sounds, I did not decide which channels to send each section to until I set up the installation and was able to realize everything. Having four sections and six channels, I was able to set up a dedicated channel for each theme as well as route duplicate sounds into another two channels as I saw fit.

Execution

For the presentation of the installation, I envisioned a bare, gallery-like space with wooden floors where the speakers could be placed around the room like a square. For the first public showing, I was able to use a room in the Palau de les Artes in Valencia, Spain. The room was very much what I had envisioned using, but a bit large. My setup included placing a table in the center of the room, which held two Mac Minis that were running my patches from Max. The computers were then running into an Apogee Ensemble in order to send out eight discrete channels. Also on the table held a stereo pair of speakers playing the music and facing opposite each other. Around the room I placed the other six speakers forming a rectangle towards the center of the room. The installation included four channels of video that I was running from Jitter and for this I set up four projectors in the room, each facing a wall.

The idea behind this setup was to have the audience experience a barrage of sounds as they entered the center of the room. After experiencing my installation first hand, I believed that I had succeeded in creating this evolving soundscape that I had desired. By adjusting the programming on every sound, every sound was being heard as frequently or as infrequently as I wished. By integrating undulating volume changes, the audience could hear bits and pieces of samples jut out of the soundscape. The sounds that I manipulated in other programs had created a wash throughout the room, which acted as an underbed for the other.

Conclusion

This was my first venture into the task of creating a large-scale multimedia installation. It was a challenging experience creating it entirely within Max and Jitter, but by using Max, it allowed me to have flexibility in programming.
Singing, the act of generating musical sounds with the human voice, requires producing exact pitch values according to the tonality and rhythmic structure of the musical piece that is sung. Most singers are capable of producing this sounds, however there are also a significant amount of people who have not the ability to sing in the correct pitch, due to the lack of proper musical training. This is the reason that AutoTune technology has become indispensability in the music industry. The term AutoTune is based on 2 main steps: Pitch detection and pitch correction. Various pitch detection and pitch shifting methods have been developed until today. In this paper, writers aimed to achieve a comprehensive analysis over these techniques that are being used in today. This paper will outline two different stages of pitch correction as Pitch Detection and Pitch Shifting. In terms of Pitch Detection; Harmonic Product Spectrum (HPS), Autocorrelation methods have been evaluated. In addition to that, an empirically calibrated fundamental frequency finding method (called finding the first peak method) has been introduced. For the subject of Pitch Shifting, Pitch Synchronous Overlap and Add (PSOLA), Phase Vocoder and Resampling methods have been evaluated.

1) Introduction

Digital signal processing techniques are widely used in music industry. Due to the simplicity of modifying audio signals, music producers often prefer using AutoTune software. The Autotune process contains two main steps: Pitch detection and pitch correction. In the pitch detection step, the spectrum of given audio signal will be obtained. In the Pitch-Correction step, the information obtained from the first step will be used to determine the shifting factor. Our approach for the AutoTune process is summarized in Figure-1.
In order to create the time-frequency representation of the signal, the given audio signal has to be divided into smaller sections. This process is called windowing, as the name states, the audio signal has to be multiplied with a window. Reasoning behind windowing is clear: Since humans are not capable of producing more than one fundamental frequency at a time, a sound signal that has been multiplied by a window, preferred to correspond to sample length of 20ms (The length of a window is important factor as frequency resolution is dependent on it. The phenomenon of frequency resolution will not be explained in this paper since it is not the aim, but reader can see [1] to get detailed information.) has to consist only one fundamental frequency in one window length. Three widely used windows were evaluated (Hanning, Hamming and Kaiser windows) and among them hanning window appeared to be the best for our benchmark audio data (C Major Scale). Secondly, to create the spectrogram of the signal, Fast Fourier Transform (FFT) is applied to every windowed data. As human voice does not represent a perfect sine wave, the spectrum of every window has to contain harmonics as well as the fundamental frequency. Determination of the fundamental frequency of every window is one of the two main steps of AutoTuning process and called Pitch-Detection process. Lastly, in the pitch correction process, the pitch of the voice is changed at any time interval to the desired frequency value.

2) Pitch Detection

Since human voice does not produce a perfect sine wave, it is not easy to extract the fundamental frequency of the sound. A sample spectrum can be seen in Figure-2. The fundamental frequency of that sound sample is 122 Hz (the first peak of the plot), in addition to fundamental component, the spectrum contains some harmonics as well.
The whole problem of pitch detection can be summarized as separation of fundamental component from the harmonics. Although a large number of methods proposed for pitch detection, authors decided to evaluate Harmonic Product Spectrum (HPS), Autocorrelation and Finding the First Peak (ftfp) methods. All methods are tested on C Major Scale. Spectrogram of C Major Scale is shown in Figure-3.

**Figure 2** Spectrum of Saygin’s attempt to create C3, fundamental frequency is 122 Hz (the first peak)

**Figure 3** Spectrogram of C Major Scale
2.1 Harmonic Product Spectrum (HPS)

HPS is based on multiplying the original spectrum with its downsampled versions, which will result a peak that corresponds to the fundamental frequency [2]. The algorithm is quite simple: Since downsampling in frequency domain will cause decimation, as the signal’s spectrum gets downsampled, every frequency component will be divided by 2. For example, if a signal that has spectral components of 100 Hz and 50 Hz gets downsampled, the new spectral will change to 50 Hz and 25 Hz. As the signal’s spectrum gets downsampled every time, the result is multiplied with the original spectrum. In the end, the highest magnitude component will be selected as fundamental frequency. Intuitively, this is sensible since this process causes the fundamental frequency to have highest magnitude value among all spectral components. The algorithm of HPS is summarized in Figure-3.

However, this algorithm has some shortcomings. First of all, it is not known that how many harmonics a certain window contains. As the downsampling process is related to number of harmonics, if number of harmonics is misestimated (so downsampling is overused) it may cause the fundamental frequency component to attenuate as it gets multiplied by near-zero elements. To show this phenomenon, C Major Scale was used to test the HPS algorithm by getting downsampled 6 times. The result is given in Figure – 5.
As it can be seen in Figure-4, HPS algorithm was quite unsuccessful to create a useful time-frequency representation of the signal. The algorithm has been run again with downsampling factor of 4 and result can be seen in

**Figure 5** C Major Scale analyzed with HPS, downsampling applied for 6 times

**Figure 6** C Major Scale analyzed with HPS, downsampling applied for 4 times
Result of HPS with downsampling factor of 4 is tremendously better and can be worked on. Consequently, downsampling amount is crucial for HPS algorithm to successfully extract the time-frequency representation of the signal.

2.2 Autocorrelation

The Autocorrelation method finds the period of signal by comparing the waveform with itself shifted by a time $\tau$. The method is well explained in [3], a simplified version will be mentioned in this paper. Autocorrelation method is different than other methods as it works on time domain instead of frequency domain. General straightforward equation of comparing two waveforms is shown in Equation-1.

\[ \text{Equation-1 General equation of autocorrelation method} \]

As the human voice is a periodic signal, when the copy is slid across by factor of $\tau$, it should match the original signal at some point. This point will be accepted as period of signal thus the fundamental frequency of sound can easily be found from it. This method is robust at mid to low frequency range; therefore it is quite popular in audio processing. However, autocorrelation has some drawbacks; it does not work well with quickly varying signals. There are some methods which increase autocorrelation method’s robustness but only the standard and most primitive version has been evaluated in this paper. The result of autocorrelation is shown in Figure-7.

\[ \text{Figure 7 C Major Scale analyzed with Autocorrelation} \]
2.3 Finding the First Peak (FTFP)

This method accepts the first peak that exceeds a pre-determined threshold as fundamental frequency. The threshold is determined by the maximum peak that spectrum contains, rather than a constant value. The general form of the algorithm is depicted in Figure – 8.

![Amplitude](image)

**Figure 8** Finding the First Peak method (FTFP) accepts the first peak that exceeds the threshold as fundamental frequency

Even though this technique gave better results than the "Harmonic Product Spectrum" and "Autocorrelation" methods, it needed some filters to be developed in order to reduce detection errors. Since the project will work on human voices, regarding the fact the characteristics and the color, namely the timbre will be depended to the human that generates the voice; it is especially hard to find constant values to build the filters upon. Instead it is more optimal to use trained values to construct the filters.

The detection algorithm uses such trained values in order to prevent inaccurate measurements. Assuming that a "normal" human voice’s frequency will vary between 100 Hz and 1800 Hz[4] a detected frequency value between this interval will be saved to a vector which will later be used for the next detection. Multiplied with 0.6 and 1.5, these values will determine the interval of the potential future detections. If the next detection fails to enter this area, an extra detection is applied to the window. The extra detection tries to locate a peak in the area determined by the trained frequency value. The possible peak will then compared with the first peak. If the second peak is beyond a threshold determined by the first peak, the second peak will be assumed to be the fundamental frequency. Otherwise the detected frequency will be equalized to zero, meaning the analyzed section consists of noise or silence. Result of FTFP method is shown in Figure-9.
3) Pitch Correction

Once the fundamental frequency is known with the help of detection algorithms, it can be used to compute how much the fundamental frequency and its harmonics should be shifted. The equal temperament scales that are often used in Western music were selected as the basis of the target pitches [5]. Any other frequency components were treated as an out-of-tune pitch since its value is not matched to any of the Western music scale notes.

The general concept of three pitch-correction methods (PSOLA, resample, phase Vocoder) will be explained in this section.

3.1.1) Pitch Synchronous Overlap and Add (PSOLA)

One of the main problems of pitch-shifting is that the duration of the sound will get affected according to the shifting amount. This change can be compensated by either adding parts of the signal to itself or by removing specific parts of the signal [6]. The add or remove process can be achieved by dividing the signal into various segments and later pushing them together or moving them apart as seen in Figure – 10.
3.1.2) Resample + SOLA

Changing the pitch of a signal by resampling is another effective method since the harmonic contents of the signal will be highly preserved. However, resampling will also change the signal’s duration. To overcome this problem, parts of the signal could be repeated or removed as seen in Figure – 11 [7].
3.1.3) Phase Vocoder

Phase Vocoder (PV) can be used to change the pitch without altering the duration and keeping the quality of the audio signal [8]. This method is widely used in the literature and mainly based on Analysis, Unwrapping and Synthesis stages.

In Analysis stage, the signal is sequenced by applying window function. Then FFT is applied to windowed data. To compensate the FFT’s frequency resolution, PV uses the phase information. For instance, if one applies a FFT of length 1024 to an audio with sample rate of 44.1 KHz, one would have 44100/1024*2 = 21.5 Hz primary frequency resolution. This means, if there were a spectral component of 35 Hz is present in the audio signal, this component will not align neither to 21.5 HZ nor to 43 Hz and would cause problems in the inverse FFT process. To resolve this problem, PV uses the phase unwrapping process.

The Phase Unwrapping (PU) stage is mainly based on finding the exact frequency components of the signal. To achieve this, PU matches the signal’s phase of each frequency bin with the previously processed window’s phase of corresponding bin. Differences in phase can be used to calculate the FFT’s bin frequencies. [9]

In the Synthesis Stage, the signal is simply stretched/narrowed to keep its duration while increasing/decreasing its frequency. After this process, the time domain audio signal can be obtained by taking the inverse FFT of synthesized data. This newly generated audio signal must be played on different sampling rate in order to keep the duration of signal and increase/decrease of signal’s frequency value. This new sampling rate is calculated with a simple operation. If one wants to increase the frequency from 300 Hz to 350 Hz of an audio with sample rate of 44.1 KHz the newly generated signal must be played on (see Figure – 12). On the other hand the signal can also be resampled by the same ratio in order to keep the same playback sampling rate.

Figure 12 General algorithm of Phase Vocoder (PV)
3.2) Performance Evaluations of the Methods

Speed and accuracy tests were run on the algorithms. All tests were made on Windows 7 with Intel I3 2.40 GHz CPU and 4 GB RAM.

3.2.1) Speed Tests

The speed test was made with a 3 seconds of audio file. The time results gathered from MATLAB’s profile summary are listed below.

- Resample + SOLA: 0.314 s
- PSOLA: 0.847 s
- Phase Vocoder: 5.172 s

Due to the complexity of the Phase Vocoder algorithm, it takes generally more time to process the audio signal.

3.2.2) Accuracy Test

The same audio test file was used for the accuracy test. The original and the modified signal’s frequency are depicted in Figure – 13, 14 and 15 for each method respectively. The test audio file contains mostly the pitch A2#. The aim was to shift it to B2.

![Figure 13](image1.png)

**Figure 13** Original Spectrum (blue dots) and shifted spectrum (red dots) of resample method

![Figure 14](image2.png)

**Figure 14** Original Spectrum (blue dots) and shifted spectrum (red dots) of PSOLA method
3.2.3) Evaluation of the Test Results

The test results showed us that, for small pitch shifting amounts, all three methods satisfied our requirements. However when high amount of shifting is required, the quality of sound gets deteriorated as it gets processed by PSOLA and Resample methods. Fortunately, the Phase Vocoder method is significantly more accurate than other methods when high amount of shifting is needed. The PSOLA and Resample methods produce echoing and reverberating that occurs as sound artifacts.

On the other hand the Phase Vocoder is computationally expensive, creating a trade-off between speed and accuracy.

To sum up, all three methods are effective according to the user’s need. As studied in the literature [10], according to the subject the most effective method could be altered.

REFERENCES

- Polikar, Robi. The Wavelet Tutorial
- Noll. A.M. *Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate*, Proceedings of the symposium on computer, 1969
- Lemmetty, Sami. Review of Speech Synthesis Technology
- Parviainen, Olli. Time and pitch scaling in audio processing
- DPW Ellis, A Phase Vocoder in Matlab, 2002
- Nikolas Borrel-Jensen, Nis Nordby Wegmann, Andreas Hjortgaard Danielsen, Real-time Pitch Correction
Music and Audio Production for Documentary Films “Over the example of a particular documentary: Yılan Hikayeleri (Snake Stories)”

CIHAN İŞIKHAN
DOKUZ EYLUL UNIVERSITY
cihan.isikhan@deu.edu.tr

Abstract

Music and audio production for documentary film have different features than the productions in other media. Making music and audio production for documentaries requires a different approach than the one used for film music and audio recording for the audio professional. Because a documentary is a whole, and only one music covers the documentary. The music is then divided into fragments which are considered suitable for different scenes of the documentary and used accordingly. Audio processing (effects, mixing etc.) comes last in the process.

All these processes are considerably time-consuming. If it were only ten years ago, these specific processes would take a lot of time. However, “the digital” changed that. Over the course of ten years, the audio world adopted the new digital audio technology which is faster to work with, easier to use, and has a higher quality.

In this paper, the process of music and audio post-production for documentaries will be described over the example of a particular documentary film called Yılan Hikayeleri (Serpent Stories). The problems that are encountered during music and audio production as both sound engineer and musician are discussed. Finally, some recommendations will be given to the young musician-engineers who are at the beginning stages of their professional life.

Introduction

At the bottom of many academic studies on music technology, lie the scientific researches and studies. Each one valuable in itself, many studies in various fields have been presented and discussed so far. Many studies, from the newly developed hardwares and techniques to softwares, from professional recording and dubbing methods to the specific areas of music technology such as music information retrieval systems, are the academic subjects of music technology presented in a wide range so far. This type of academic studies in general, as well as in other disciplines, prepare a theoretical background to the field of music technology. Theoretical background triggers other studies and creates a cascade of motion process. Thus, the applicable size of the science that is the absolute reality would be revealed. The interactional and sustainable application fields of music technology including sound recording, dubbing, publishing, music arrangements, etc., are the application fields at the very final point of such an academic process which is initiated academically and finalized with the current demands and naturally with the human factor.
Also, outside the academic process, regardless of any theoretical ground, there are studies that are directly based on applications. These studies can be roughly expressed with various words, such as popular, disposable, quick and random. These are such expressions that, even if one rolls up their sleeves to form a scientific study basis on this subject, the study would barely end up in the mid-pages of a popular journal. Because the resultant has not gone through a process, it is ordinary, and, let’s name it; “it’s completely a story”.

However, this is a very wrong perspective. Owsinski insists that a process in music technology which is described academically as above, makes sense only when it is matched with the “ordinary” (Owsinski 4). In fact, according to the author, application fields which are regarded as ordinary, have such value that will make way for the academic studies in music technology.

It should be confessed that, we, the music technology academics in Turkey have made the same mistake in many scientific studies so far. In the papers and articles, either the specific was investigated or the thoughts were conveyed. Sometimes it was focused on a novel technique or a software, and sometimes aesthetic concerns were brought to the foreground. However, none of them had presented the direct application, the one which was previously defined as "ordinary". Until the presentation by Can Karadoğan (Karadoğan 81-87)...

When Karadoğan chose the subject of his communique as the communication /interaction between the sound engineer and the artist in productions, we thought that this was completely a convection of application, and so it was briefly and roughly conveying the "ordinary". However, the results did not leave such an impact at all. Contrary to expectations, the importance of a direct application convection, i.e., what is popular and actual in academic level was understood well and clearly.

With the urge to pick up contextually where Karadoğan’s observational studies have left off, a study coming from an application will be presented in this study. For the application field, music and sound effects production process for the documentary entitled Yılan Hikayeleri (Serpent Stories) which the production period lasted approximately for two years and has been completed recently was chosen. All the the process from the first presentation of the documentary for music and sound effects to me by the director to the delivery will be strived to be explained in this communique. Therefore, for the reasons explained above, this paper contains an entirety which actually, with Owinsky’s words, is expected to show direction to the academic studies and therefore reflects a direct application, rather than being a presentation of an academic study carried out in the field of music technology.

The Documentary "Serpent Stories" and the Plot

One thing that all music technology workers agree on, which is confirmed when considered historically, is that the initial point of music or, with a more general expression, sound technology dates back to visual technology, briefly to movie and video technologies. Apart from Echo recording his voice to the mountains in the tale by the Latin poet Ovidius, Plato’s mechanical musical instrument which works with water and is said to play a melody in every hour at the 4th Century BC, and the Monk Magnus’s machine which produces human voice at the 13th Century AD, the beginning of music and sound technology dates back to the late 19th Century, to Martinville and Edison. Recording and listening practices that Edison introduced to the world with phonograph, is actually
Nothing more than film production and exhibition curiosity which starts with Martinville's sound recordings and is introduced to the daily life with Edison's kinetoscope... (Ünlü 27). The late 18th century which are the birth years of the film industry, as expected, is also the beginning of the film productions with documentary purposes. However, it is not possible to talk about a professional production until the documentary called Moana which is about Samoa Island and released in 1926 (Ellis, and Betsy 23). Documentary is actually referred to shoots for military purposes rather than the popular one. Therefore, front shoots and others made during the years of World War I are the first actual documentaries known. After 1960, in both technological and general terms, the so-called contemporary documentary productions begin. Departments of government organizations for documentary purposes only, like the BBC in particular, continue documentary production today with all their amenities.

It would not be a mistake to say that documentary music has no difference from film music in terms of production, and that they only have a visual difference in terms of their durations.

The documentary film Serpent Stories which is the subject of this study, tells the semi legendary semi true stories of rural population living in the Aegean Region with the snakes. The documentary starts with consecutive telling of different stories, and with the contributions by archaeologist and art historian scientists, reveals how the tales that date back to old ages and today's stories match up. These stories which gather the past and the present, sometimes directly coincide, whereas sometimes undergo various changes and evolve even if the subjects stay the same. In fact, as in the example of snake as a symbol of medicine, these changes show how wrong even some world wide accepted facts are. Because, the two entwined snakes figure which is the symbol of contemporary medicine today, is actually not the symbol of medicine in the past but it represents a publishing house that prints books on medicine.

Main Music

Every producer in the music production does and should have a style. My style for this documentary and all the documentaries alike, is producing a single orchestral music that will cover the entire documentary that lasts approximately for 4 minutes, instead of music production unexceptionally for every scene. This music, with the contemporary words, can be described as a "single soundtrack" for the documentary. It's a music with a clear introduction, development and finale parts, formed around a relatively catchy melody, instrumental, and even if not necessarily, in "new age" genre. Seeing the entire documentary may not be necessary for this kind of music. I can say that, a brief information given by the director even at the beginning or during the shoots on the outlines of the content of the documentary is sufficient for me for a production in such style. Because the whole documentary is important, not the scene.

In this study, I proceeded to such a production method. When the subject of the documentary was explained to me and a theoretical fiction was offered to me that can be considered as a guide, both the theoretical and practical music production process lasted approximately for 2 months. The resultant was the orchestral "main theme" that lasts approximately for 4 minutes which can easily be named as Serpent Stories.
The main theme was produced with expectations of belonging directly to the documentary, intending to evoke the documentary even without seeing it, and always reminding of the documentary again when it is listened to, after watching the documentary. This music was directly the music of the documentary, but on the other hand, this music was actually a guide for the documentary. The main theme was presented at the introduction, at the most striking scene and at the finale of the documentary. The melody was recurred in some scenes as a reminder.

**Scene Music and Silences**

Every documentary has a visual main theme. However this theme consists of a combination of a number of side-themes which are conceived as independently as possible from one another. We can briefly call these side-themes "scenes". These visually designed scenes naturally should have music of their own. These music should be designed with independent genre and arrangements from each other and each one should be independent from one another. The scene should start with the music of its own and the music should end again with the scene. The important thing here is the transition between scenes creating a whole with the music. The scenes and music that start with the ending of another is called "cut roll" and the scenes and music designed with a transition between each other are called "a/b roll". Therefore, designing the scene music as "cut" and "a/b" is an important matter for the whole documentary.

There are a total of 11 scenes in *Yılan Hikayeleri* and the same number of music is produced for every scene. These scenes are, respectively:

- Introduction
- Generic Speeches
- Expert Speech Introduction
- Stories 1
- Expert Speech 1 (Serpent in Mythology)
- Stories 2
- Expert Speech 2 (Serpent in Sociology)
- Stories 3 (animated narratives)
- Expert Speech 3 (Serpent in Theology)
- Stories 4 (Serpent in Islam)
- Finale

When the music produced so far was placed consecutively, they are sorted as the main themes and the scene themes with their visuals. At first glance, all this traffic may have a continuous feeling of the music resonated throughout the documentary, from the beginning to the end. However, with John Cage's words "silence is also music". In this regard, continuous visuality does not leave a negative impact on the audience, even the gaps in the visuals may create a big problem, however, and the situation is different in music. The continuity of music, although not yet certain, may create a negative effect on the documentary whole.
The scenes which have no music, is usually the scenes that contain effective speeches or speechless visuals even as effective as speeches. However, the silence of music is generally seen in parts with the ambient speeches (not in perforations). In case of frequent existence of such conversations in the documentary, the silence is preferred in hardly heard speeches due to shooting circumstances or in the scenes with contextually the most striking speeches.

In fact, the music documentary is completed within the process explained so far. The main theme was used in the introduction and the finale and recurred in notable scenes. The scene music was aligned with the independent visuals throughout the documentary and at other times, silence (no music) was used due to reasons explained above. Thus, all the music for the documentary was produced.

**Sound Effects**

As in every visual product, the most important sound source after the music production were the sound effects. A wide range of effects were produced, for Yılan Hikayeleri, from various animal sounds, primarily snake, to nature sounds, from movements of humans to artificial sounds such as guns and etc. For all the sound effects, libraries produced by sound effects producers which are expert in their fields (BBC, Boom, SoundScape, etc.) were used. For example, for the dry firing and gun burst sounds used in a story impersonation, provided that the same brand of gun, a variety of sound effects were overhauled, and among them, a wave file produced by Boom with a sampling rate of 192 kHz was chosen. Effects were chosen according to the related scenes with same diligence (never downloaded from the Internet). As a result, the documentary was edited with sound effects all benefited from libraries.

**Mixing and Mastering**

The whole music was produced VST based in Cubase and mixed in ProTools. However the mix of the documentary containing the effects was carried out in Audition. The biggest reason for that was that the documentary will be edited in Premiere. Using Audition maintained a healthier swapping of the sounds that are conceived by me between me and the final editing operator.

The music production was done entirely by using virtual libraries. However, cello playing the melody of the main theme was recorded live in the studio. The biggest reason for that, I would say, was that even if the virtual cannot be distinguished from the real, the virtual instruments become increasingly artificial particularly in the mixing stage and gradually differ from its real counterpart. Therefore, an exception was created for cello which plays the melody as the striking point of the documentary, and hence, the main theme.

Another exception was also was true for the percussion. Because the content was a documentary and therefore the visual impact became prominent, the more cinematic sounding instruments became also more prominent. Of course, these issues may be open to debate, but I can say that the percussion has the most cinematic impact among all instruments and the best library that is giving this characteristic in virtual environment.
is Hans Zimmer Percussion by Spitfireaudio. Therefore, all percussions formed around the main theme were selected from this set and this created a special situation for the percussions with the cello in this documentary.

Following the completion of the recording phase completely, Cubase Mix Automation was utilized in the mixing phase. Also, primarily, Waves, Izotope and Melda Mixing automation softwares were used for mixing purposes and plug-ins of these same softwares were used for the mastering.

Conclusion

One of the most important working fields of the music technology is the music arrangement editing. Especially, considering the current technology, this field has created a specific field of its own. The industry that this field directly serves is undoubtedly the music industry, however, it wouldn't be wrong to say that it has an equivalent flow of service to the visual arts, prominently to movies and documentary genres.

Within the context of visual arts, documentary genre has a more special place in music production, compared to the movies. With their shorter and more certain durations, representation of visual flow with the content in a shorter time and their necessity to describe more in less time, documentaries have a special place in visual arts. This difference in documentary production is undoubtedly reflected to the music production. Music for documentaries, in general terms, are short and clear. They facilitate instantaneous transitions. It contains an entity, although there is no such obligation.

The documentary *Yılan Hikayeleri* and its music explained in this study was prepared within this context. However, in an unconventional way in music production, a music theme that lasts approximately for four minutes was produced for this documentary, and this music was either used directly at the introduction, finale and the most striking two scenes of the documentary or the melody of the theme music was played with a specific instrument. For the rest of the scenes in the documentary, different music was used with different alignings, and although few in number, only some sound effects were used without music in some scenes.

The music was entirely produced in computer environment by using virtual instruments, only a live recording of a cello was made for the melody of the main theme. For the virtual instruments, a wide variety of libraries including Vienna Symphonic, Hans Zimmer Percussion and Voxos were utilized. All the recordings were mixed and mastered by also using virtual hardwares, all the visuals and audio were combined during editing.

Of course, several challenges have also emerged all throughout this process. The most prominent of these challenges is the urge for "re-editing according to the finalized music after the completed edit", which is a common characteristic of all directors. Apart from these, some director demands such as "the delivery deadline for this documentary was yesterday" or "I was feeling different previously from what I want now, let’s change it" and etc., were also experienced in this production process.

Technically the biggest problem was experienced during the mixing of the virtual instruments. Virtual instruments resonate realistic alone, but they become increasingly artificial as they are combined with other virtual instruments. Therefore, the most challenging phase of the production was the mixing. This problem can be solved by using a detailed processor as much as possible and recording the instrument live (as in the cello) when necessary.
References


The dividing lines between the different stages of music and post-production are blurred in today’s fast-paced conditions. These days the steps like recording, post-production, mixing may proceed concurrently, overlapping with each other. If I try to analyse the current state of a typical post-production, I could summarize it as:

- The inception and the design of the idea take place almost simultaneously.
- To edit and organise the project, we no longer wait until all of the recordings are completed.
- Thanks to the computing power of current computers, creatives are able to work remotely, detached from dedicated and expensive hardware in the studios.
- Mixing may start even at the beginning, or early stages of the production. Again, thanks to the computing power, mobility and softwares, there’s no reason to wait.

The filming of a typical weekly TV series in Turkey takes about 4 days. The editor/director prepares the 1st rough cut on the day 4 or 5, and then the film is sent to audio post and music producers, although sometimes the final cut continues to be prepared and evolved almost until the day the show airs on TV.

The tools, technology and the techniques we use are the definitive factors on how well we perform and keep up with the fast pace in these conditions, and it’s not always easy to keep sane...
The current state and difficulties in Turkey

If we are asked: "What are the ways and principles that are used in Turkey?" it wouldn’t be easy to answer. Because it’s hard to see any efficient workflows and it usually looks like a big mess. In the countries where the film industry is an established one, there also are established methods or ways of handling different production steps. One of the main reasons for this, is the necessity and the means. Developed countries are also the ones which invent things and create the needed technology.

The other thing is the approach and mentality. Creative people all around the World usually require more time to be creative but the tools and the people use those tools don’t have the luxury to procrastinate and be slow – if the workflow is not a fast one, someone will have to lose a lot of time in a dark room that doesn’t have a lot of fresh air. Time is money, and someone has to spend one or the other.

The other thing is the quality of work, or the product. There will always be a relation between the money spent and the quality of work wherever we are, but usually quality suffers because of the inefficient ways of doing things. Reasons might be the skimpy production companies that are hiring wrong people – that needs to be researched.

We lose a lot of valuable time in the studios in Turkey. The same requirements apply in the USA, UK and India too - everything should be finished and delivered fast. But compared to developed countries, Turkey does not suffer because Turkish artists are less creative, but because we Turks don’t use the existing technology properly. There’s no real collaboration between the video and audio professionals because they don’t speak the same language when it comes to project interoperability. Could we use that time we lose in the process elsewhere? Definitely.

The use of technology, project/file formats and the efficiency
I think one of the major reasons why our workflow suffers is the technical incompetency. There are not many people knowing and using the required software and file formats, which could facilitate handshake between the video editing and audio creation softwares. The way we share the files and projects is another issue as well.

Avid Media Composer and Apple Final Cut Pro are the most popular Non-Linear Video Editing softwares among the professionals in the USA, while Adobe Premiere and Edius are the most popular ones here in Turkey. Adobe Premiere and Edius are popular among Indie or hobby filmmakers in the developed countries. The choice of the video editing software indirectly leads to faster workflows, because producing and releasing an Indie film is a lot different than producing comparatively bigger budget commercial films. When things get more serious, the team working on the project gets bigger and it becomes more complex to manage everything.

Although the favorite software of choice of Turkish video editors like Adobe Premiere is able to output a file/project format that audio professionals could easily use, the missing part is communication and training. If we add the piracy issues -as in buying a software is an unpopular behaviour among hobbyists- one could only guess that these shortcomings is rooted from the past of the individual. Editors might be carrying on their habits from their early years and not learning more during their professional work.

There are project formats used in order to enable video people work with audio people. The most popular ones are being OMF and AAF. AAF is more advanced and have more features (and it is aptly named: "Advanced Authoring Format") compared to OMF.

For an editor that uses Media Composer (by Avid) or Final Cut Pro (by Apple), exporting the project in OMF or AAF format is almost an everyday task.

The USA is considered to be the heart of film industry and as we all would agree, everything should flow fast and efficiently in order to be at the top of the business.

Since the digitization of the film industry started in late 80s, Avid has been the frontrunner in the Non-Linear Video Editing (NLE) technology. Although the competition is fierce now with the addition of Apple Inc. with its Final Cut Pro to the equation, Avid Media Composer is still the leader in the business.

On the audio front, there had been another leader in its own field, Digidesign with its audio hardware offerings and Pro Tools software attached to them. It became an instant hit within the commercial studios in the US at the beginning of 90s and regarded as the “Industry Standard”. Maybe to get a better hand in the competition, Avid bought Digidesign and added Pro Tools into its product line.

1 The most popular Video NLE softwares among semi-pro editors in the US: Adobe Premiere, Apple Final Cut Pro, Sony Vegas Pro according to (http://www.learningdslrvideo.com/nle-market-share-breakdown)
The marriage between the Avid Media Composer and Avid Pro Tools has created a familiar environment for many production companies and it lead many producers to choose these products over the competition.

It's not hard to guess that these two softwares can create compatible project files to each other not only by using OMF and AAF authoring formats, but also creating and exchanging proper META data which makes it easy to manage complex working sessions. Especially within Media Composer version 8 and Pro Tools 11, there are advanced integration features speeding up the workflow.

"But wait, there’s more"

Editing video and audio in a non-linear fashion has one disadvantage... it spoils the creatives and adds additional layer of complexity to the work of operators and editors, mainly because the decision taking could last forever and kill the operator during the process.

The TV series in Turkey are usually shot as 100-minute films, which are regarded as feature length movies in the US. Now imagine you have to finish a 100-minute feature film every week, for the next 13-something weeks minimum. Including video producers, audio post producers too, should work around the clock and turn in their work in 6-8 hours. Both the creation and the delivery of the mix becomes a hard labor in this case.

Adding to this, the time needed to deliver the physical medium between the producers has to be managed and calculated. The files are delivered in DVD-Rs and HDDs, and usually a guy with a motorbike may knock your door around 4 AM in the morning to give them to you. Or take something from you (Hopefully just the files, nothing more, brr). Well, forget about the new Internet file transfer services available but now... not even FTP?

Revisions – The Unavoidable

Music editing in Pro Tools - from one of the episodes of the TV series "Mihrap Yerinde"
Everyone knows the time is limited, so the director/editor of the film sends the video files to re-recording engineers, dialog editors, composers, music producers as soon as s/he has the rough cut.

Working on an unfinished film is pretty scary, especially if all you do is timing your cues (music) tied to the picture. Even though editing in an audio editor or DAW software is a Non-Linear process just like the video editing software, with the changes in the scenes, everything will have to be shifted in the timeline of the film. If you don’t have a method to re-align your work with the future revisions automatically, you WILL lose at least 1/4th of your studio time while trying to catch your own tail.

And that’s when the EDL and Conforming comes into the picture.

Modern Workflows – EDL and Conforming

While the revisions start knocking your door, there are only a couple of things that could ease your mind: EDLs and Conforming. If the video editors can and be willing to supply them to you of course, otherwise your best bet would be coffee and the vitamins (you should quit smoking if you haven't done it already). EDLs are "Edit Decision List"s and should go together with the edited video, preferably in the same file folder for ease. With the help of EDLs and conforming software, it would be possible to automate the adjustment of your work to the revised picture. The conforming software (sold seperately) "reads" the changes in the timeline, and edits the cuts that previously made to their new locations in the time line.

Now that is what we are supposed to use in Turkey in video, music, and audio post-production but we don’t, yet. This process could save tremendous amount of time and energy of many operators and could lead to a better quality of work.

Edit-Change Lists could also be used with other software than Avid’s and Apple’s – the video NLEs like Adobe Premiere can export, and Audio DAWs like Nuendo can import EDLs and conformed by the conforming softwares. All these tools are available and they are cross-platform.
Conforming Softwares

Virtual Katy (commercial product) - Windows, OS X, Pro Tools
Ediload (commercial product) - OS X, Windows, Pro Tools,
The Maisch Conformer (commercial product) - Windows, Pro Tools, Nuendo, Cubase

Thanks to these techniques and softwares, there’s no need to look at the revisions with disgust.
But there are more improvements that can, and has to be made.
We humans always push the boundaries of what we can achieve with what we have and our abilities, and that pushes the technology and invention.
While the technology enables us to go faster and maybe to the undiscovered territories, we need to utilise our tools in the way that –at least- previous century allows.
We have and we should use the Internet to our advantage to make our work and ourselves independent from physical boundaries, buildings and countries. But of course, we need a new mindset to be ready for the true revolution and follow the inventions closely. Or just, follow the white rabbit...

Conforming Softwares

Let alone the operators, producers, and artists, even the audience is enjoying the benefits of cloud computing (maybe sometimes without noticing) in their mobile phones and tablets. The resistance is futile.
The media companies and content creators should adapt themselves otherwise they will lose business. If not now, it’s going to happen soon.
In every aspect of production, whatever the industry is, the technologies that can automate the hardwork will replace the human labor and will let us creative humans deal with the creativity. We even have the tools like WaveRider plugin, which creates an automatic volume automation pass for the mixing engineer to review and adjust later.

We already use file servers at the facilities that our Workstations reside to get connected and use the files from. We call them "pool"s. There's nothing wrong with that system and it will exist as long as desktop computers exist, but it's not as open as the cloud servers – you need to be in the same network all the time at your facility. Almost every musician and sound designer has a private studio now and the industry is moving away from big facilities because they are difficult to manage. There are more and more small teams working on big productions because of their flexibility and creativity.

The next step is and will be virtualisation and using Internet services to speed up the collaboration and delivery mechanisms. We audio post and music producers output big files that won't fit into e-mails but can send them via the Internet thanks to fast, broadband speeds and clever way of sharing the files.

We see more and more companies investing on the Cloud business and there are already established companies specifically serving music and media producers. Gobbler is one of them. Gobbler differentiated itself from the competition (i.e. Dropbox) by creating specific features for audio and video files like special file size compression algorithms.

Gobbler application analyses the project folders, understands the structure and relationship between the audio and session files, compresses the audio into FLAC format on the client side before uploading, then uploads them to the cloud for sharing, backup and archival purposes.
While a typical consumer sees the cloud services as mere storage units that are accessible online 24/7, the things we could do with them is much more than that. We’ve already seen the collaborative potential of Dropbox, mainly because this service has a big user base, which creates more showcases than Gobbler can. For now.

Actually, the services like Dropbox had become wildly popular among “bedroom musicians” and they were the early adopters in this regard. The first time I used Dropbox was when my Norwegian musician friends wanted me to collaborate with them on their Eurovision Song Contest submission back in 2010. I was already subscribed to the service, but was using it just as an online backup and file server for my files. But this time we shared the project folders and were using Cubase as the main DAW on that song.

Now there are online music collaboration services like blend.io, which incorporates Dropbox as the cloud storage service and makes the online collaboration a breeze.

The users login to their accounts and browse for the like-minded musicians across the globe (or at the same town – location does not matter) and see if there’s a music track that they would like to work on.

Although Avid supported Gobbler at the beginning by opening its Software Development Kit and putting a menu within Pro Tools 11 software dedicated to Gobbler for easy upload and syncing, after deciding that Avid too, should have its own Cloud service called “Avid Everywhere”, they parted their ways.

Then Gobbler hired the key staff from Avid to its own team, things got heated up and Avid sued Gobbler.

And this case shows how some people and companies (ie. “Industry”) see a great potential on this “Cloud thingy”.

While a typical consumer sees the cloud services as mere storage units that are accessible online 24/7, the things we could do with them is much more than that. We’ve already seen the collaborative potential of Dropbox, mainly because this service has a big user base, which creates more showcases than Gobbler can. For now.

Actually, the services like Dropbox had become wildly popular among “bedroom musicians” and they were the early adopters in this regard. The first time I used Dropbox was when my Norwegian musician friends wanted me to collaborate with them on their Eurovision Song Contest submission back in 2010. I was already subscribed to the service, but was using it just as an online backup and file server for my files. But this time we shared the project folders and were using Cubase as the main DAW on that song.

Now there are online music collaboration services like blend.io, which incorporates Dropbox as the cloud storage service and makes the online collaboration a breeze.

The users login to their accounts and browse for the like-minded musicians across the globe (or at the same town – location does not matter) and see if there’s a music track that they would like to work on.

Or you could create a private project and collaborate with only the ones you want to work with – it depends on your preference.

Once you get the hang of it, you realize how quickly things could turn very interesting and people may work on your song while you are sleeping.

In the screenshot on the left, you see me “pulling” the Pro Tools session of the track ”Saints” by Moby – another artist who is a bedroom-musician-turned-Global-artist in the early ‘00s.
You are encouraged to work, and do a remix if you want too – no need to deal with his manager and ask for it after jumping through hoops and making expensive telephone calls.

In blend.io, the user allows write access to a new, specific folder in his/her Dropbox account dedicated to the service and the files get transferred from cloud to cloud in just seconds. Mainly because, there’s nothing to copy, they were already there, just "assigned" to a different account.

When a user wants to copy a file from one folder to another in his/her local computer and hard drive, it really gets copied and can be seen in 2 different locations. No, it doesn’t have to be that way on the Cloud. It’s clever, and usually safer than your own drive.
But of course, we need more assurances on the privacy and security of our files and projects when it comes to the films and sensitive artistic projects. There are a couple of shortcomings of Dropbox application currently. The complete privacy of the files is not guaranteed with additional layer of encryption during file transfers, and the folder hierarchy of Dropbox is a limiting one.

When the video and audio post producers work on their Workstations, some of the files are kept in certain separate folders than the working directory of the project. And we can’t connect and sync those "outside" folders yet.

We are yet to see a working product version of “Avid Everywhere”, but from the demonstrations we see, it’s clear that it will be a much more comprehensive platform than the likes of Dropbox and Gobbler.

But for now, Gobbler seems to be fitting into a few bills and workflows.

Other than the additional functions of Gobbler (like file type-specific compression algorithms before transfer), if one needs a better version of Dropbox, s/he could consider the service from SugarSync.

It lets the user pick the folder(s) of choice; it has comprehensive folder-based permission settings, and seems to be having more serious file security policy than Dropbox. Cheaper subscription plans doesn’t hurt either. (And I also see Avid Training Administrators using SugarSync for sharing the education materials with Education Partners across the globe).
All of these tools and services are there to be discovered, used and be inspired; enabling us to work more efficiently and fun. I truly feel we are lucky to have these and I think it only can get better.

I just wish that my colleagues and counterparts on the video side would too, start using the same or similar tools so we all could benefit from the new age of technology. Let the technology work for us – that’s the whole point of having the technology anyway, isn’t it?

Presenting the Examples:
"Hadi Baba Gene Yap" (Feature film) Dialog editing, sound design, mix
"Mihrap Yerinde" (TV Series) Music, mix
Softwares used: Pro Tools 11, Nuendo 6.5, WaveRider 3 (plug-in)

References
http://www.learningdslrvideo.com/nle-market-share-breakdown
http://resources.avid.com/SupportFiles/attach/EDLManagerGuide.pdf
http://stevencastle.wordpress.com/2013/04/24/pro-tools-conformer/
http://www.maisch.at/TMC/TheMaischConformer.html
In this letter, the linear predictive coding (LPC) method is studied to generate spectrograms. Linear predictive coding generally has been applied in speech-recognition systems for many years. This work aims to apply the LPC method to audio signals and also distinguish the main differences between the traditional methods and the LPC method. Short time Fourier Transform (STFT) is the most common method employed to generate spectrograms for signals. Performance of the STFT has been proved in many applications. In this work, linear predictive coding method’s computational steps and the comparison with the short time Fourier Transform ARE examined.

I. Introduction

The ability to identify the patterns in speech and audio signals has a great importance in recognition systems. In particular, for audio applications, differentiating various patterns will give crucial information about the signals. In speech processing, there are different methods to extract the information from speech signals. Linear predictive coding is one of these methods which has been well understood for many years and it has various applications areas.

In this work, basics of the LPC are employed to extract information from audio signals. Moreover, the mathematical details and computational process is briefly explained. Signals can be accurately modelled using the LPC coefficients. In voiced regions, the LPC behaves less effective. However the model still includes important information on the signal. The LPC model is a simple approach due to its linearity. Moreover, it is also easier to implement compared to the traditional feature extraction methods.

II. The LPC Model

The basic logic behind the LPC model is that a given speech or audio signal at time of \( n \), \( s(n) \) can be associated with a linear combination of the past \( p \) speech or audio samples:

\[
(1)
\]

where the \( a_1, a_2, a_3, ..., a_p \) are the assumed constants and the \( s(n) \) is the represented signal at the time of \( n \). Equation (1) can be easily translated to an equality by including an excitation term, \( G u(n) \):
where $u(n)$ is the normalized excitation and $G$ is the gain of the excitation. If Equation (2) is expressed in $z$ domain:

(3)

Then Equation (3) leads to the following transfer function:

(4)

Equation (4) output is defined as $S(z)$ and the input is defined as $G U(z)$.

Figure 1 Linear Prediction Model of Speech [1]

The interpretation of the Equation (4) is given in Figure 1, which shows the excitation and Gain parameters are multiplied and then inserted as the input to the all pole system to produce the speech or audio signal, $s(n)$.

Figure 2 Speech and Audio synthesis model based on LPC model [1].
The actual excitation function for speech is either a quasiperiodic pulse train (for voiced speech sounds) or a random noise source (for unvoiced sounds). The valid LPC model is presented in Figure 2. In this figure the normalized excitation speech is chosen by a switch to differentiate the voiced and unvoiced sounds. Then, the gain of the source is estimated by the speech signal and the scaled source is used as an input to a time varying digital filter which is controlled by the vocal tract parameters [1]. The output of the time varying digital filter is the characteristic of the speech signal.

Another equation that is derived from the LPC model is the prediction error, $e(n)$, defined as:

$$
(5)
$$

Where $u(n)$ is defined as the predict signal and the $s(n)$ is defined as the original signal. It can be easily extracted using Equation (4):

$$
(6)
$$

The prediction error, $e(n)$, will be equal to $G u(n)$ if $s(n)$ is actually generated by a linear system such as the one presented in Figure 1.

The basic problem of the linear prediction is to determine a set of coefficients which is defined the original signal in time and the spectral properties of the digital filter, shown in Figure 2., match those of the speech waveform within the analysis window.

There are several methods to extract linear prediction coefficients in a signal. Autocorrelation and covariance methods are both used to extract linear prediction coefficients. In this work, LPC coefficients are found by using the autocorrelation method.

I. Introduction

![LPC Processor Block Diagram](image)

**Figure 3** LPC Processor Block Diagram

The ability to identify the patterns in speech and audio signals has a great importance in recognition systems.
Figure 3 shows the block diagram of the LPC processor. Analyzed speech signal belongs to a male and contains three words ("It didn't work"). Signal’s sampling frequency is 22050 Hz and signal’s time is 1.5 seconds. The block diagram is including the following steps;

1. Pre-Emphasis Filter
   The pre-emphasis filter is used to spectrally flatten the signal and make it less susceptible in the signal processing. The most widely used pre-emphasis filter is the fixed first order system:

   \[(7)\]

   Where \( a \) is the filter coefficient. Difference equation can be written as:

   \[(8)\]

   In this work, pre-emphasis coefficients are selected 0.97 for this particular example.

2. Frame Blocking
   Pre-emphasised signal is blocked into frames of \( N \) samples, with an overlap number of \( M \). In speech processing the number of overlaps is generally chosen by the given equation which is defined as \( M=1/3N \) [1]. Frame blocking is applied to the system since the adjacent frames overlap and the resulting LPC spectral estimates are correlated from frame to frame. If \( M \) \( N \), then the LPC spectral estimates from frame to frame will be quite smooth. However, the some parts of the signal will be lost if \( M \) is higher than \( N \) and the correlation between the resulting LPC spectral estimates of adjacent frames will contain noise. In this work, \( M \) is selected 300 and \( N \) is calculated 100 based on the relation provided previously.

3. Windowing
   To minimize the discontinuities within the signal and taper the signal to zero at the beginning and the end of each frame a hamming window is used after the frame blocking. If window is defined as \( w(n) \),

   \[(8)\]

   The basic window used for the autocorrelation method of LPC and the method used in most of the recognition systems, is the Hamming window is defined as:

   \[(9)\]

   Effects of the hamming window into the framed signal are presented in Figure 4.
4. Autocorrelation Analysis

Each frame of the windowed signal is next auto-correlated to obtain:

\[(10)\]

where the highest autocorrelation value, \( p \), is the order of the LPC system.

5. LPC Analysis

The next step is the LPC processing, which converts each frame of \( p+1 \) to autocorrelations into the *LPC Parameter set* which the set is the LPC coefficients. The formal method for converting from autocorrelation coefficients to the LPC parameter set is known as the Durbin's algorithm. The details of this algorithm are not discussed in this work. Durbin's algorithm result provides the LPC parameter set.

6. LPC Synthesis

LPC synthesis is used to reconstruct a signal from the error signal and the transfer function of the signal. Because of the signal transfer function is estimated from the LPC coefficients these coefficients can be used with the error signal to construct the original signal. An example of the original signal, reconstruct signal and the difference between the original signal and estimated signal which is called error ratio, can be seen in Figure 5. For this example 15 LPC coefficients are used. It can be easily seen that the error is relatively small to prove the original signal is presented well by the LPC estimated signal.
7. LPC Spectrogram
The constructed spectrograms based on the LPC with the 20 coefficients is compared with the traditional STFT in Figure 6. In Figure 6, it can be seen that the LPC results are more analogous compared to the STFT. Moreover, LPC spectrogram is more generalized than the STFT. LPC spectrogram is averaged out some components to represent the signal with a lesser set. This is a disadvantage of the LPC spectrogram. On the other hand, LPC coefficients represent the signal in a fine model which consists the powerful frequency components.

IV. Results
In this work, another method to generate spectrograms is examined. LPC model is proved that its results are useable for generating spectrograms.

Reference
There seems nothing new in singling out the eternal return of the theme of repetition in the critical history of Wong Kar-wai’s *In the Mood for Love* (2000) and his oeuvre as a self-referential whole. In her Sight and Sound review published in 2000, Amy Taubin notes the dreamlike quality of the repeated shots of Maggie Cheung climbing the stairs in her sensuous cheongsams. Rey Chow in “Sentimental Returns: On the Uses of the Everyday in the Recent Films of Zhang Yimou and Wong Kar-Wai” argues that the repetition of everyday objects (qipaos and shoes worn by Maggie Cheung, the rice cooker, the thermos, the radio, newspapers) and daily locales (offices and restaurants, deserted street corners, and interiors of taxicabs and households) serves to “conjure a subjective, yet pervasive, mood of melancholy.” In “We Won’t Be Like Them: Repetition Compulsion in Wong Kar-Wai’s In the Mood for Love,” Nancy Blake observes “a quality of visual loop” manifested in the “various repetitions, doublings, and symmetries” in the film, as if “the film reel has skipped backward or performed the visual equivalent of an old LP getting stuck in a groove.” Peter Brunette points out that the repetitiveness of the couple’s street encounters is achieved through a constant change of cheongsam. Stephen Teo notes two running motifs set up by repetition in *In the Mood for Love*: first, the motif of role-playing; second, the motif of change. Liu Yunghao in his article “Rehearsal, Repetition and the Illusion of Love: Analyzing Wong Kar-Wai’s *In the Mood for Love*” perceptively shows the connection between repetition, rehearsal, and retakes. David Bordwell observes cyclical scenes such as “routines of cooking and gift exchange, hesitant pauses, sheer down-time” from other Wong films and repetitions of “locales, musical cues, camera setups, and floating slow motion” in *In the Mood for Love*. Jean Ma observes “a discourse of doubling wherein motifs and themes reappear from story to story.” Another way to put this is to say that *Chungking Express* and *Fallen Angels* should be thought of as a diptych. Likewise, the “mode of intertextuality” interpenetrates *In the Mood for Love* and *2046*. Repetitions operate not only from film to film but also within each film. Take *2046* for instance, Jean Ma nails down “a logic of repetition that governs the construction of character in *2046*” in the doubling or mirroring of Chow Mo-wan in Tak and, by extension, of Wong Kar-wai in Chow Mo-wan.
Repetition as a musical concept

While the foregoing scholarship on In the Mood for Love acknowledges the centrality of repetition in structuring the film, evoking a sense of nostalgia and past-ness of time, registering the everyday and the change within the repetitive, few film scholars have approached repetition as fundamentally a musical concept and realized that any attempt to make sense of its use within/beyond the film should not bypass the question of musical repetition. The earliest scholarly writing on music's contribution to the charm of Wong Kar-wai's oeuvre is Emelie Yeh Yueh-yu's "A Life of Its Own: Musical Discourses in Wong Kar-Wai's Films." She maintains that music is the key to unlock the "mysteriously hypnotic quality of Wong's style" because "music performs a discursive function in his work," that is, music has semantic meaning in his work. However, instead of delivering what her title promises—a study in musical signification or music semiotics—Yeh's approach betrays at once a limited understanding of music as song and a linguistic bias toward discourse analysis of the lyrics; hence her preference for popular songs over instrumental music, emotive particularity over ambiguous expressivity, and the representational over the non-referring. What I mean by Yeh's "limited understanding of music as song" has to be distinguished from Raymond Monelle's idea of music as "wordless song." Whereas Yeh sees an unbridgeable gulf between meaningful vocal music and vague instrumental music, Monelle regards the music of song as "no less absolute than that of an instrumental piece." In other words, we can talk about the semantic dimension of the music of song as well as the lyrics of song: "music is already song, before any text is added." So when Yeh restricts the production of meaning primarily to lyrics and secondarily to musical genre and arrangement, Monelle maintains that musical meaning exceeds the boundaries of textual references of the lyrics and sometimes exists in a confrontational relationship with the text.

"A Life of Its Own" was written one year before In the Mood for Love. Nine years later in 2008, when Yeh finally had a chance to tackle the "transcultural sounds" in In the Mood for Love, she focuses only on Zhou Xuan's "Full Blossom" (Hua Yang De Nian Hua), a 1940s Mandarin pop, overlooking sonic aspects of Zhou Xuan's rendition as well as other orchestral music in the film and making no reference to an excellent master's thesis from the Graduate Institute of Musicology at National Taiwan University titled "The Transformation of Rhythmic Visual Images — the Interactive Relation of Music and Image in In the Mood for Love," which meticulously tracked the beat-to-beat melodic movement of the most repeated non-diegetic music "Yumeji’s Theme" and the second most repeated song "Quizas, Quizas, Quizas" in relation to camera movements, cuts, and character movements. Macedo de Carvalho in her article "Memories of Sound and Light: Musical Discourse in the Films of Wong Kar-Wai" regards the repetition of "Yumeji’s Theme" as embodying "the indetermination of the lovers as they repeatedly approach and retreat from each other: The music appears a total of eight times throughout the film: first when the characters meet; twice when they go to the noodle stall; twice when they are together in the hotel room; twice when they are isolated but thinking about each other; and one last time when they decide to go their separate ways." Despite their stimulating insights into the film, Yeh and Macedo de Carvalho fail to take stock of the intersections of the pre-existing main theme and the film and the latter even ignores there is a ninth repetition of the main theme playing while the end credits are rolling.

A whopping nine repetitions of Shigeru Umebayashi’s "Yumeji’s Theme" are
distributed throughout Wong Kar-Wai’s 2000 film In the Mood for Love, drenching the period piece in the mood of at once anachronistic and dislocated nostalgic lyricism. By "anachronistic and dislocated" I mean the failure of "Yumeji’s Theme” alone to evoke the film’s temporal and spatial setting, which is Hong Kong in the 1960s. In fact, listening to "Yumeji’s Theme” with its attendant visuals tells us neither its name nor its origin—the obsessively recurring theme is not so much "of 1960s" as extracted from the underscoring of Seijun Suzuki’s 1991 film Yumeji, the third of the director’s Taisho trilogy. The use of pre-existing "Yumeji’s Theme" and its persistent repetitions thus pose a question of poetics: What does it mean for "Yumeji’s Theme” to repeat nine times throughout In the Mood for Love? Or to pose the question differently from the perspective of audience perception: How does the audience perceive and experience the nine repetitions of "Yumeji’s Theme” when they audio-view In the Mood for Love?

These two questions of authorial poetics and audience perception become even more problematic when the issues of sonic realism and historical authenticity are involved. Wong Kar-wai claims that he attempts to recreate not only the authentic "look" but also the historical soundscape of the bygone era. On the visual level, along with the ravishing cinematography contributed by Christopher Doyle and Mark Lee Ping Bin, the production designer William Chang Suk-ping crafts the sensuous, tight-fitting cheongsams enveloping Maggie Cheung’s elegant body and equally elegant gait. On the aural level, based on his personal experience of moving from Shanghai to Hong Kong in 1963 when he was five years old, Wong Kar-wai conceives of the soundscape of In the Mood for Love as representative of Hong Kong "radio days,” comprising traditional operatic excerpts (Peking opera, Cantonese opera, Shaoxing opera) and modern popular songs (Spanish, Mandarin, and English). Set against the "authentic” production design and the reconstructed historical soundscape, "Yumeji’s Theme” can claim neither sonic realism nor historical authenticity yet it occupies most space in the musical landscapes of In the Mood for Love. My initial two questions "What does it mean for 'Yumeji’s Theme' to repeat nine times?” and "How does the audience perceive and experience the nine repetitions” are compounded by a third one: How does "Yumeji’s Theme” figure in relation to the historical soundscape, costumes, and production design in the film? I propose two paths, theoretical and pragmatic, to unpack these questions. The theoretical path is to revisit the Stravinsky’s famed denigration of film music as aural wallpaper via Peter Kivy and Ben Winters while the pragmatic one is to conceptualize the praxis of sound design through the mind of Wong Kar-wai’s sound designer Tu Du-Chih in hopes of grappling with what I call "the idea of intertexturality” operating simultaneously on macro and micro levels of repetition and recontextualization in In the Mood for Love.

Igor Stravinsky: from musical wallpaper to aural fabric

In 1946, a provocative article "Igor Stravinsky on Film Music” was published in two parts in Musical Digest, documenting a conversation between Igor Stravinsky (1882–1971) and the composer Ingolf Dahl (1912–70). In his polemic against attributing any artistic values to film music, Stravinsky claims that "[t]here are no musical problems in the film. And there is only one real function of film music—namely, to feed the composer!" If we look past Stravinsky’s sarcastic censure against film music’s commercial orientations, we can see that his real target is the utilitarian, extra-musical ends of film music. That’s why Stravinsky stresses that “[m]usic explains nothing; music underlines nothing.” That is, film music lacks narratological functions and expresses nothing
emotional and realistic references. Stravinsky's agenda is the aesthetics of absolute music, which he puts in rather categorical terms: "good music must be heard by and for itself, and not with the crutch of any visual medium." The sire of neo-classical style denounces music's subservience to the moving images because the "visual crutch" cripples the supposed "self-sufficiency of music" and distracts the audience from the true contemplation of music proper.

Paradoxically, due to his bias toward architecturally and structurally rational absolute music, Stravinsky provides revealing insight into an underrated decorative function of film music—its wallpaper function. Film music is to "bridge holes" and "fill the emptiness of the screen" just like wallpaper covering the empty spaces of walls, which spells bad music for Stravinsky because wallpaper music is by definition background music and operates subliminally on the audience. Thus, film music cannot be heard "by and for itself" in active contemplation of its play of musical form and has to be considered side by side with the visuals. Stravinsky’s inadvertent insight doesn’t stop at the visual and tactile wallpaper metaphor. He further elaborates: "The orchestral sounds in films, then, would be like a perfume which is indefinable there." The olfactory aspect is added to the wallpaper metaphor. In the conclusion of the article, Stravinsky sets himself, the composer creating divine order and rational organization through music, against "the cinematic concept of music," which utilizes music "like remembrances, like odors, like perfumes which evoke remembrances" "for reasons of sentiment."

I want to subvert Stravinsky’s hostile use of the musical wallpaper metaphor by literalizing it and arguing for, rather than against, audio-viewing film music as aural fabric constituting photographic as well as phonographic texture of the film. While Stravinsky denigrates film music’s multisensory and affective appeals, attributes that contradict his cult of absolute music, I hear in In the Mood for Love a radically and sensuously evocative use of film music, and absolute at that in Raymond Monelle’s sense of music as "wordless song," being foregrounded through repetition and sheer volume in combination with visual slow motion, and thus can no longer be relegated to the subliminal perceptual background and demand active and immersive listening. Stravinsky’s synesthetic and multisensory wallpaper metaphor can serve as a theoretical prism through which we come to conceptualize "Yumeji’s Theme" and its excessive repetitions. The idea of audio-viewing the musical repetitions as aural fabric in In the Mood for Love has to be apprehended through multiple sensory modalities—vision, hearing, touch, the sense of temperature—that constitute a multidimensional orientation. Likewise, the musical wallpaper is to be perceived stereophonically rather than stereoscopically, in multidimensional rather than two-dimensional terms. Let’s now move from Stravinsky’s take on film music to music philosopher Peter Kivy’s philosophizing on the function of musical repeats in the context of absolute music.

Peter Kivy: from sonic wallpaper, decorative art, to sound design

Stravinsky’s verdict on film music as wallpaper has struck a surprising chord with synesthetic and multidimensional conception of film music as aural fabric. The wallpaper metaphor serves to shed light on questions of poetics, perception, and placement of musical cues. It asks the ontological question of what film music is and the phenomenological question of how film music is heard and perceived through other senses and combined with the visual medium. Now I want to invoke Peter Kivy’s theorization
of musical repetition in the context of absolute music, that is, music without text and without extra-musical function, and apply it to the context of the musical repetitions of "Yumeji's Theme" throughout In the Mood for Love. In so doing, I'm strategically hearing the whole film as a piece of musical composition and the literal repetition of "Yumeji's Theme," albeit of different lengths according to the rhythm and pace of editing, as musical repeats of a certain section.

Kivy teases out three competing models—the "literary" model, the "organism" model, and the "wallpaper" model—to conceptualize the nature of absolute music. The "literary" model, with its agenda of ascribing discursive meaning to instrumental music, cannot fully explain the function of musical repeats because language prefers "continuous flow" to repeating and returning to "previously mentioned" material, let alone repeated verbatim nine times. Neither can the "organism" model, with its teleological directionality of progress and development, account for the seemingly "redundant" musical repetition. Tracing Kant and Hanslick's philosophical lineage of placing music among the decorative arts as "sonic decoration," Kivy proposes a "sonic wallpaper" or "sonic carpet" model to account for the ornamental pattern of musical repetition without making any reference to Stravinsky's wallpaper conception of film music, which is understandable given that Kivy's interest lies in absolute music rather than utilitarian music. At the same time, Kivy preempts the charge of trivializing music by setting sonic wallpaper apart from its visual counterpart. Sonic wallpaper, Kivy argues, has four distinguishing features: it is multidimensional, quasi-syntactical, deeply expressive, and deeply moving.

For me, the strength of Kivy's sonic wallpaper theory lies in his situating the function of musical repetitions in the perception of recurring pattern in the context of decorative design. Musical repetitions, Kivy asserts, "both the external ones, in which whole sections of works are literally played again, or internal ones, where small musical figures or patterns are reiterated, are the means by which the composer of the sonic carpet makes his design, in the large and in detail." If we strategically stretch Kivy's sonic wallpaper theory a bit by substituting the composer with the filmmaker and the sound designer, it can gesture toward a theory of sound design in film. In the Mood for Love can be seen and heard as sonic wallpaper or sonic carpet in which sections of "Yumeji's Theme" are literally replayed again and again and precisely placed in filmic texture. The repetitions of "Yumeji's Theme," to quote Kivy again, "[compel] us to linger; to retrace our steps so that we can fix the fleeting sonic pattern." When we move from sonic wallpaper theory to a theory of sound design as symbiotic with visual design, the wallpaper theory can accommodate a filmic acoustemology beyond the confines of absolute music, acoustically knowing and making sense of the relationship between music, sound effects, and tone of voice in the sound design of In the Mood for Love.

**Ben Winters: from musical wallpaper to sonic environment**

In his article "Musical Wallpaper?: Towards an Appreciation of Non-narrating Music in Film," Ben Winters moves away from Claudia Gorbman's narratological model of film music toward Carolyn Abbate's concept of film music as "unscrolling" rather than "narrating." Furthermore, Winters revisits Stravinsky's wallpaper metaphor and argues for the use of film music's expressive qualities to shape sonic environment. The idea of sonic environment within film narrative should be related to both our multidimensional understanding of "Yumeji's Theme" as aural fabric and our invocation of the placement of
musical cues as part of sound design and production design. This is particularly interesting in the case of In the Mood for Love and Wong Kar-wai’s modus operandi—treating production design, costume design, and editing as a holistic artistic trinity embodied in his longtime collaborator William Chang Suk-ping. Their collaboration started since Wong’s directorial debut As Tears Go By (1988). Chang has designed and edited all of Wong’s films from Days of Being Wild (1990), Chungking Express (1994), Ashes of Time (1994), Fallen Angels (1995), Happy Together (1997), to In the Mood for Love (2000), 2046 (2004), My Blueberry Nights (2007), and The Grandmasters (2013). William Chang once says during interview: "I cannot separate costumes and sets because for me, those two things are coherent. I like to control the costumes and the sets together, and I have done so since the beginning." xvi I contend that sonic environment is part and parcel of the costumes-and-sets bundle William Chang talks about and sound design should be considered along with production design. The fact that William Chang doesn’t take up the responsibility for sound design is, I conjecture, less because of aesthetic concern than because of technical contingency.

When we audio-view the nine repetitions of "Yumeji’s Theme" as "wallpaper music' and the traditional operatic excerpts and snippets of popular songs of the era as "sonic spices," the placement of musical cues becomes ultimately weaving of "aural fabric" with the visual sequence, the matching of the aural texture with the textural details of Maggie Cheung’s tight-fitting cheongsams, the advertisements and cracks and marks on the walls, the surface of staircases and pavement. Wong Kar-wai’s working with his production designer William Chang and sound designer Tu Du-Chih can thus be conceptualized as what I call tentatively an aesthetics of intertexturality in sound design, that is, the intimate interweaving of musical cues, sound effects, and human voice with the moving image, the close textural analysis thereof can only be addressed in the longer version of present paper.

Resources


---

i "The shot is repeated at least three times, each repetition accompanied by the same slow dissonant mazurka on the soundtrack. The music, the slo-mo, and the incongruity of the elegant dress and the clumsy rice bucket make the moment seem like a dream." Amy Taubin, "In the Mood for Love," *Sight and Sound* 10.11 (2000). I should point out that the accompaniment is waltz rather than mazurka. Both in triple meter, the waltz has the accent on beat 1 whereas the characteristic mazurka rhythms displace the accent to the weak beats (second or third beats) of the bar.


ix Ma, "Chance Encounters and Compulsive Returns," 136.


xi All Stravinsky quotes in this section, unless indicated otherwise, are from Mervyn Cooke, "Igor Stravinsky on Film Music (1946)," *The Hollywood Film Music Reader*, ed. Mervyn Cooke (New York: Oxford University Press, 2010).

xii One could argue that the figure and ground perceived in wallpaper make it more than two-dimensional.


Indie Games: A Case Study

MASSIMO AVANTAGGIATO

CONSERVATORIO G. VERDI MILANO

mavantag@yahoo.it

Abstract

In this paper we're going to illustrate the way graphics and audio are combined in Seeds of Death, a videogame produced by Italian indie game developers. Seed of Death is a science fiction horror-themed FPS (first-person shooter) game: the player leads the role of a space marine fighting his way through hordes of zombies and other invading enemies.

This game uses a prototype of Oculus VR technologies, a virtual reality head set for 3d games: we talk about immersive gameplay, game peripherals integration and virtual reality boundaries.

Our aim is to describe how making music or sound design implies an in-depth knowledge of the world of games as well as the ability to integrate graphic engines (Unity) and middleware software (FMOD).

1) An Indie Game Production

When we talk about indie developers, we refer to "small development teams with somewhat larger commercial interests. Though still working on relatively small budgets, these teams produce games with the intention of creating a product available for sale to customers through various channels[1]". Independent developers "retain operational control over their companies, pipelines, and organizations and often may work with proprietary engines or other proprietary software". Indie game developers are part of large communities; they often try to cooperate through social networks and forums.

These small budgets teams are placed at an intermediate level between hobbyist and larger game publishing companies.

Seed of Death is a small budget game with no creative constrictions, very common for indie productions, but aims at reaching the qualitative results of larger development teams.

The aim of this ambitious Italian project was to create a product available on line and sold in official stores in 2015.

---


3 Game will be launched with New Oculus prototype called Crescent Bay: the new headset features 360-degree tracking and it is a much lighter model that integrates headphones and make virtual reality more comfortable and immersive. http://www.theverge.com/2014/9/20/6660901/oculus-announces-new-vr-headset-prototype-crescent-bay.
2- Why Combining FMOD and Unity?

Most commercial games with considerable sound design use live mixing and sound management approach based on Middleware Audio Software: FMOD, Wwise, other custom tools.

Considering the high cost of licences, these tools were used in the past almost exclusively by large companies; nowadays middleware companies, such as Firelight Technologies, the company behind FMOD, have changed considerably their commercial policies and are trying to capture the market of Indie game developers.

In September 2014 Firelight Technologies announced that tools will be free not only for non-commercial projects but also for commercial indie projects. This new trend is explained by Firelight Technologies CEO Brett Paterson:

We started out as an indie company so we’ve always felt an affinity with the indie community. They’re risk takers, innovators and explorers. We wanted to give back to that community and we can’t wait to see what they do with FMOD Studio.

Middlewares “...help teams meet deadlines, lower production costs and can help elevate the results of any project - large or small”.  

But... Why combining Unity and Fmod? Unity doesn’t provide any mixing tools and sound designers will probably go insane if they have to manage a lot of sound objects in the Unity editors.

Even if every developer should take into account FMOD limitations, this middleware software allows composers and sound designers to leave old developing pipelines for new workflows. Particularly, composers can leave old Linear Audio Production (LAP), in favour of Generative Audio Production (GAP), with the aim to create better quality sounds.

Unity has had a large diffusion, because it offers some remarkable advantages:

1) Unity asset store can speed up graphic and game development: this is an important plus for small development teams;
2) Unity offers quick-to-make prototypes: it is much easier to make simple 3D games than in other 3d platforms.
3) Visual prototyping offers a large series of benefits;
4) Unity has a large community of users, also between indie game developers.

---

4 Interview to Martin Wilkes (Firelight Technologies):
http://www.gamedev.net/page/resources/_/business/interviews/interview-with-firelight-technologies-r2776
A negative influence of audio is often caused by repetitive or boring sound and music, unpleasant effects, wrong type of music, unrealistic and unconvincing sounds.

An important advantage of a Middleware Audio solution is that it enables sound designers to edit, mix and master game audio using the appropriate interface while the game is running. This allows to appreciate the impact of sounds and music on player’s immersion and to avoid unpleasant influences of audio.

Particularly, the approach of Generative Audio Production, described in FMOD Manual\(^5\) allows us to:

1) work more efficiently, separating the activity of the composer, that creating events, and the activity of programmers and designers setting up the "event triggers"\(^6\):
2) to reduce the time of implementation;
3) to reduce resources usage: each of the sound components can be reused in other projects

---


\(^6\) “A trigger is an element that is called by the game and then defines a specific response in middleware to accommodate what is happening in the game. More specifically, in interactive music a trigger responds to a spontaneous occurrence in the game and launches a stinger. The stinger, which is a brief musical phrase that is superimposed and mixed over the currently playing music, is a musical reaction to the game. [...] The game would call the trigger which in turn would launch the stinger and your music clip would play over the ongoing score. The game engine triggers an event to change the ambient sounds playing in a game”. Wwise Reference
3-Composition, Microcomposition and Horizontal/Vertical Interactivity

Horizontal interactivity in music is a sort of composing technique where instead of writing one long theme, a composer creates multiple, smaller phrases which can musically connect and interchange with one another. In this way, composers can create music that responds dynamically to game events. A composer is also required to organise the musical material in such a way that alternative sequences of the fragments are supported.

This kind of horizontal re-sequencing has various advantages such as a limited requirement on data storage and a reasonably fast responsiveness.

In Level One we studied several variations of the same theme: the longer we explore, the more variations will randomly play.

The music theme "Deadly Moves" in the Third Level (the Spaceship) is based on game state: music changes depending on whether the main character explores, fights or reaches a new game zone. In this case the piece moves from one phrase of music to another in a linear fashion in order to make the system react to certain events.

In other cases e.g. when we meet a final boss, music themes' change is based on the player's health conditions: this contributes to create a dramatic suspense and to increase the emotional involvement of the player.

In general terms, we tried to stimulate the player emotionally and we have allowed more musical diversity by 'cueing' up multiple possibilities: F MOD Designer's visual composition tool allows the creation and visualisation of complex branching structures.

In certain cases, horizontal interactivity in music is based on a 'micro-composing' technique. Many sound elements, that respond dynamically to game events, such as explosions, effects or soundscapes, were created through a "microcompositional" technique. E.g.: Some background ambiences in the Third Level (Spaceship) were built using a combination of short metal effects; shrill metals; metallic stingers whoosh; singing bowls sounds; metallic synth drones.

Generally speaking, in music the horizontal approaches are combined with vertical techniques, "creating a sophisticated musical mix which can adapt in intensity, phrase variation, as well as thematic variation". For example, in the orchestral piece "Intermezzo Sinfonico", each orchestral section is grouped into 'stems'; each stem could be played alone or in conjunction with any other stem. E.g.: music switches to a new situation when you encounter the boss or when you defeat him.

---

8. It also allow the composer to create musical transitions In Seeds of Death musical transitions are based on: looping; key change; crossfading; moving from one musical genre to another; using one-shot phrases to overlay important moments (stingers); using filters.
9. FMOD allow to create explosion sound effects that uses generative audio: "The end result will be a sound effect that is unique each time it is triggered and utilizes control parameters to vary the output sound depending on the desired material content of the explosive event". FMOD Studio User Manual, p. 224.
4- I.E.Z.A. Model: Labelling Sounds

This conceptual framework [4] was developed between 2003 and 2008: it offers a model for understanding the relationships between game audio and game immersion [5] [6]. For this reason this model has been widely used during the game development sessions: all the sounds, generated using different commercial and non-commercial tools, were classified using the categories of the I.E.Z.A. model and referring to the game level in which they were used.

This model uses two dimensions to describe sound in games:
1) the first dimension makes a distinction between the diegetic and extradiegetic level [7].
2) the second dimension, which is more relevant to us, considers domains or categories: Interface, Effect, Zone, Affect.
Affect and interface describe a non-diegetic level [8]; zone originates within the diegetic level.
1) Interface: expresses what is happening in the game: menu option buttons, information about the gameplay, information about health, power-ups, other game status signals; status bars, pop-up menus and score points.
2) Effect: describes what is happening in the game world.
E.g. First Level (hospital): footsteps, dialogues, weapon sounds, punches, main character’s body movement sounds; the approaching of enemies (zombies’ hoardes; spiders; final bosses).
3) Zone: describes the geographical, cultural and topological setting of the game world.
E.g. First Level (hospital): Hospital Blood Pressure Tester; Hospital Ventilator Compressor; Hospital Ventilator running; Hospital heart monitor; Hospital Nurses Station Voices; Slow/medium/Fast squeaking; Helicopter Army Approach and Pass; Doors Creaking.
4) Affect: reflects the emotional status of the game or it anticipates upcoming events in the game: it originates within the diegetic level.
E.g. First Level (hospital): various sound effects such as risers/wooshes; various horror soundscapes.

5- An Immersive Experience?

We can define immersion as an emotional involvement that consists in a strong connection with the game: the player is highly concentrated on the objectives of the game and has slow reaction to the real world.

According to Huiberts (2010, 45), even if there are many different definitions of immersive experience, three basic aspects are present in most of the available classifications:
- transportation in the game world, absorption in the game activity, identification with game situation.

---

11 “Diegetic sounds originates from within the fictional universe, where we assume that the characters perceive the sound as natural to that world, while extradiegetic sound is a commenting aesthetic feature with no direct connection to an actual source within the fictional universe, and consequently, the characters do not seem to be able to hear it. Jorgensen, Kristine (2009, p.98).
In Seeds of Death, all these three aspects have been considered during implementation. The game, in fact, offers a real immersive experience thanks to:

a) Immersive soundscapes, generative audio effects and music.
b) A careful Artificial Intelligence Implementation: the behaviour of Non-Playable Characters, is sometimes unpredictable;
c) The use of optional peripherals such as Oculus Rift and Razer Hydra: these game peripherals allow the exploration of the boundaries of virtual reality.

The final version of the game will support two players together over a network, sharing a playspace, and moving in 360 degrees with positional tracking;
d) Through the implementation of some simple narrative features — qualities: 3d intros, many video trailers between levels, high-quality speeches;
e) Some programming and technological escamotages: programmers contrived some tricks to increase the dimension of the game area through Oculus rift glasses. This has contributed to improve the state of connection to the game.

a) The suitability of the music and the non-predictability of some sound effects is an important constituent for immersion in games.

Another important aspect is the simulation of real world: effects sound often "mimics the realistic behaviour of sound in the real world". In Seeds of Death the game audio is "dynamically" processed using techniques such as real-time volume changes, panning, filtering and acoustics. Other escamotages were:

- the use of audio weather module, a plug in object that produce procedurally generated weather sound effects;
- the emulation of 3d distance: the volume of sounds was automatically adjusted to simulate distance from sound sources. e.g.: we created voices and sounds of demons to be positioned anywhere within 3 dimensional space;
- the creation of other phenomena such as generative explosions;
- the creation of alternate, multiple versions of the same sound object

We have mentioned only few techniques to make the sound more interesting.

The final objective, as we mentioned before, is to avoid any unpleasant and negative influence from audio. If the game makes you feel less immerse, sound designers have to understand why: this is a relevant issue for audio designers.

b) Much attention has been paid by programmers to the behaviour of Non-Playable Characters (NPCs). Artificial Intelligence (AI) adds behaviour to characters, that can be low or high-level: low-level behaviour such as going from a starting point to a target position; High-level behaviour such as decision making: high level behaviour adds realism to human (or human-like) NPCs.

In the field of gaming and computer science, I.A. refers to the intelligence of the machines and the effort to re-create human-like intelligence in Non-Playable Characters (NPCs). In Seeds of Death the system has a mixture of NPCs with advanced A.I. and NPCs with very poor A.I.. This mixture is wisely balanced through the level of games: for example zombies’s Artificial Intelligence is increased in advanced level.
c) Seeds of Death uses Oculus VR technologies\(^{19}\), a virtual reality head set for 3d games, that allow\(^{19}\) the first truly immersive virtual reality experience for video games\(^{20}\). The Oculus Software Development Kit (SDK) includes engine integrations for Unity 4 that make getting started with VR game development easier.

*Seed of Death* uses this kit to immerse players in a survival horror environment and to explore ambiences with peripheral vision and head tracking, and use Razer Hydra controllers to use shoot guns, build defences, pick up items.

Oculus, which combines three different kind of sensors (accelerometer; magnetometers; Gyroscopes), makes that game far more compelling than it would be otherwise; it adds to the feeling that you are really in the game and are able to go where ever you want.

A gyroscope sensor built in the Oculus provides in fact the player with the option of rotating the avatar's facing direction by rotating the peripheral. This provides a deeper sense of immersion in audio games than it would in "traditional" games.

---

12 "some genres of game are heavily dependent on narrative. Adventure games are almost entirely narrative-driven, and platformers and First Person Shooters (FPSs)". http://www.gamasutra.com/view/feature/1838/book_excerpt_and_review_game_.php
14 "Understanding the relationship between the 3D Min and 3D Max distances is essential for creating 3D audio environments. These two properties define a major aspect of how audio functions in 3D spaces. There are other properties that also influence sound behavior but they are usually closely related to the 3D Min / Max properties". Fmod Studio User Manual, p.174
16 A non-player character (NPC), sometimes known as a non-person character or non-playable character, is any character that is not controlled by a player. In electronic games, this usually means a character controlled by the computer through artificial intelligence.
All these features can fit to all kind of gameplayers, from mainstream player to hardcore gamer: there "seems to be a hint that mainstream developers are getting more experimental with interface and story/character immersion". 

**Figure 7** a little programming "trick" to increase the dimension of the area game through oculus rift. This contributes to improve the state of connection to the game.

**Figure 8** Oculus Rift integration: on the left you can see the game object hierarchies; on the right you can see some details of Oculus VR Libraries.

All these features can fit to all kind of gameplayers, from mainstream player to hardcore gamer: there "seems to be a hint that mainstream developers are getting more experimental with interface and story/character immersion".

### 7- Conclusions

A typical game project can have thousands of sound, music, and motion assets: Middleware Software allow to manage projects and assets efficiently.

Before making definitive decisions about music and sound design, the team group was provided with several versions of the FMOD Studio Project:

1) Different platforms;
2) Different localization/languages;
3) Different audio version e.g. version with music, but no sounds; version with sounds but no music; version with both sounds and music.

All the versions could be accessed by the main initial menu in the demo version. This allowed programmers to have a clear idea of how the sound design worked in the game.

[21](http://forum.audiogames.net/viewtopic.php?id=9410)
The use of FMOD has ensured that all audio is produced in line with the common, agreed audio vision for the project, creating the detailed specification of all audio elements for creation and implementation, identifying areas requiring innovation.

As the game industry continues to search for more pervasive experiences, sound productions through middleware softwares offer new possibilities that have only marginally been investigated; the use of middlewares, if appropriate, can lead to innovation and suggesting new creative solutions (Grimshaw, p.123).

References


Yip Milo, “Middleware in Game Development”, Europe-China Workshop on E-Learning and Games, Edutainment, 2005.

OTHER REFERENCES

3Dgamers Forum: maint.ign.com/3dgamers;
Gamer Forum: www.gamers-forum.com;
Indie Vault Forum: http://www.indievault.it;
Insidegamer Forum: www.insidegamer.nl;
Indiegame Forum: http://indiegames.com/index.html;
The Tweakers Forum: gathering.tweakers.net;
Music technology, involving all the technological steps during music production process, is a developing field in a close relationship with the disciplines as physics, electronics and computer sciences. From a worldwide perspective, it is obviously clear that music technology is an improving field, where new tools and software are developed every day, widening horizon for the contemporary electronic musician.

Open source microcontroller based development environments as Arduino and Raspberry Pi, arised parallel to the rapid developments in today’s computer and electronic technologies, has started to take place of predecessor microcontroller based technologies as PIC and BASIC_stamp, which requires a certain level of understanding in electronics and low-level computer programming. These tools also have the ability to convert the analog data, sensed from the outer world, like movement, pressure, light, sound, acceleration and bending of materials into useful information, meaningful for digital applications.

These new hardware, has also expedited the rise of "do-it-yourself" philosophy, which relies on self-sufficiency principle, without feeling the need for any third party expertise. These devices are also suitable for developing control interfaces for computer-based electronic music applications, by providing flexible and easy-to-implement design opportunities.

In this study, the possibilities and opportunities provided by these new generation tools and sensor devices are discussed with given schematics and code examples.

**Keywords:** Do-it-yourself, electronic music, open source, arduino

1. **Introduction**

The term *Do-it-yourself* often describes the methods of building, repairing and modifying something by someone’s own, without the help of experts or professionals. The term, which became popular by the end of the 1950s as a part of the definition of suburban husbanding, is mostly associated with home improvement and maintenance activities (Gelber 67). Today, the term is used for almost any self-sufficient activities, which do not require any assistance or help of a professional.
For computer sciences and electronics, the accessibility of "digital fabrication" and embedded computation provides many opportunities and enables anyone to produce their own electronic devices. It’s possible to design a circuit board on the computer, make the board at home and assemble the components. However, this process might be challenging for some individuals. Applications like "making of the PCB", and "soldering components" may require a certain level of experience and some components may also be unavailable for some DIYers (Mellis 22).

The open source idea, first emerged as the open source software movement, refers to a computer program, in which the source code should be distributed together with the program itself. The open source hardware term, derived from the same open source idea, also refers to any hardware design in which the mechanical drawings, schematics, PCB layout data and the software that drives the hardware are released along with the product.

Unlike open-source software, of which Linux is usually the often-cited example, open-source hardware seeks collaboration where physical objects are the outcome. It engages a distributed model of hardware development with contributors generally residing in different parts of the world. Rather than closed systems, open source projects allow an individual freedom to access the source files of a design, make improvements, and redistribute these improvements to a larger community (Evans 5).

Today, development boards as Raspberry Pi, Arduino, Beaglebone and Pcduinio provides the DIYers with many opportunities, which the developers can easily reach any documentation needed and build their own projects with a reasonable price. These devices are also capable of sensing the outer world by the help of various sensors, enabling the developers to utilize the physical events for controlling specific hardware and software.

Almost all modern electronic music equipment is controllable via MIDI or OSC (open sound control) messages. These messages are simply digital communication protocols, which makes it possible for different brands and models to communicate with each other. Using modern microcontroller based equipment and sensors, it is possible to generate these messages from sensed gestures for musical performances and transmit them to the relevant musical instrument or software.

2. Raspberry Pi

The Raspberry Pi is a small credit-card sized computer created by the non-profit Raspberry Pi foundation in the UK. The primary purpose of the designers was to bring back the simple programming environment which was widespread among kids in the 1980s on home computers such as Spectrum and Commodore 64, in order to use it for the educational purposes. Finally, Raspberry Pi first released around February 2012 with a relative cheap price of 35$ (Sjogelid 7).

Pi is equipped with an ARM family microprocessor and capable of running a GNU/Linux operating system installed in an SDcard. Like a typical computer the user can communicate the Raspberry Pi by connecting it to a computer monitor over HDMI and using a USB keyboard and a mouse. The GPIO pins that exist on the board also enable the Pi to interface with other pieces of hardware.
3. Arduino

Arduino is a microcontroller based development platform paired with an intuitive programming language that one can program using the Arduino integrated development environment. By equipping an Arduino with components as sensors, shields (special modules), lights and speakers, it is possible to turn it into a programmable brain for any control system. Because the Arduino is an open source hardware, all the design files, schematics, and source code are freely available to everybody. This advantage of Arduino makes it an easily available and popular microcontroller based developing environment for many DIYers. (Blum xxiv).

The early Arduino boards were developed using 8-bit Atmel AVR RISC (reduced instruction set computer) microcontrollers. The first board was based on the Atmega8, which had a clock speed of 16 mHz with 8 KB flash memory. The following versions used Atmega168 with 16 KB flash memory, and today’s recent versions as Arduino Uno and Duemilanove are built using Atmega328 microcontroller with a flash memory of 32 KB. For projects requiring more I/O and memory, there are also boards as Arduino Mega1280 with 128 KB memory and Arduino Mega2560 with a memory of 256 KB. (Evans, Noble and Hochenbaum 5).

The available Arduino versions today can be listed as (Arduino.cc):

- Arduino Uno
- Arduino Due
- Arduino Leonardo
- Arduino Yun
- Arduino Micro
- Arduino Robot
- Arduino Esplora
- Arduino Mega ADK
- Arduino Ethernet
- Arduino Mega 2560
- Arduino Mini
- Lilypad Arduino USB
- Lilypad Arduino Simple
- Lilypad Arduino Simplesnap
- Lilypad Arduino
- Arduino Nano
• Arduino Pro Mini
• Arduino Pro
• Arduino Fio

3.1. Arduino Uno

Arduino Uno became one of the most popular microcontroller based development boards since its first announcement on September 25, 2010. The major difference between Uno and its previous versions was the inclusion of Atmega8U2 microcontroller, which enables Arduino to be programmed over the universal serial bus (USB). By the means of this microcontroller, the board is easily recognized by the computer as a USB device, such as a mouse or a keyboard.

The Arduino Uno, equipped with an 16 mHz Atmega328 microcontroller, have 14 digital pins, each of which can be set as either an input or output, and six analog inputs. In addition, six of the digital pins can be programmed to provide a pulse width modulation (PWM) analog output. A variety of communication protocols are also available, including serial, serial peripheral interface bus (SPI), and I2C/ TWI. Included on each board as standard features are an in-circuit serial programming (ICSP) header and reset button (Evans, Noble, and Hochenbaum 5).

The Arduino Uno board is equipped with ADC’s (analog to digital converters) which can measure the voltage on the analog pins and convert it into digital numbers. These 10 bit converters are capable of measuring analog voltages between 0-5 volts and quantize this range into $2^{10} = 1024$ ($0 – 1023$) values (Blum 44).

The instructions are given to the Arduino board by programming it, using the Arduino integrated development environment (IDE). Arduino IDE is a piece of software where the developer will write the program codes (also called as sketches), compile and transfer them from the computer to the Arduino board (Olsson 15).

The programming language used within the IDE is called Arduino C. Arduino C follows a set of rules that govern syntax and structure and tell the interface board what to do. Some of these rules are inherited from C and others have been put into place by the Arduino development team to simplify usage for beginners (Evans 22). A simple Arduino sketch consists of two blocks, named `setup()` and `loop()` as shown below:
void setup() {
   // put your setup code here, to run once:
}

void loop() {
   // put your main code here, to run repeatedly:
}

Simply, any code that needs to be run once on the startup should be placed in setup() block and the code needs to be run repeatedly should be placed inside the loop() block.

3.2. Generating and Transmitting MIDI Messages Using Arduino

On an Arduino board, the digital pins 0 and 1 are used to receive and transmit serial data.

Figure 3 MIDI Out Connection With Arduino

Once the connection in Figure 3 is done, it is simple to generate and transmit any kind of MIDI messages through the MIDI-Out port. The example piece of code below shows how to generate a note-on message and send it over serial.

void noteOn(int cmd, int pitch, int velocity) {
   Serial.write(cmd);
   Serial.write(pitch);
   Serial.write(velocity);
}
It is also important to set the serial transmit rate to 31250 bauds, which is the standart data rate for MIDI, before sending the MIDI data. Otherwise, the sent messages will be meaningless for the receiving device or software. Serial data transfer rate is defined in the setup() block of the sketch as shown below:

```cpp
void setup() {
    Serial.begin(31250);
}
```

### 3.3. Reading Analog Pins

An arduino board can read analog voltage values from its analog pins and convert it to digital numeric values. This process is called analog-to-digital conversion (ADC). Analog signals can be described as continuous voltage changes, whereas, digital signals are discrete values. During the conversion process the accuracy is determined by the resolution of the system. For Arduino Uno, the resolution is 10 bits, which means that the range between minimum and maximum analog voltage values read from the analog pins (default 0-5 Volts) can be subdivided (quantized) into $2^{10} = 1024$ different values.

The example code below shows how to read analog voltage values from the analog pin 2, and print the digital values on the serial monitor included with Arduino IDE.

```cpp
void setup() {
    // initialize serial communication at 9600 bits per second:
    Serial.begin(9600);
}

void loop() {
    // read the input on analog pin 2:
    int sensorValue = analogRead(A2);
    // print out the value you read:
    Serial.println(sensorValue);
    delay(1);  // delay in between reads for stability
}
```
Potentiometers are not the only components that can get connected to the analog pins of an Arduino board. There are many types of sensors (Figure 5) that can measure real world physical quantities as temperature, acceleration, pressure, force and etc. Most of these sensors generate analog voltage values depending on their measurements. This feature of those components makes them perfect tools for translating physical actions such as body gestures into meaningful data for musical expressions.

3.4. Generating MIDI Control Messages Using Sensor Values

Figure 6 shows the connection of a MIDI out port and an IR proximity sensor to an Arduino Uno. The proximity sensor measures any obstacle's distance and generates an analog voltage value according to the distance measured.

The example code below shows how to read an IR sensor and send the measurement as a MIDI control change message:
void setup() {
    // Declare and Initialize the variables:
    int sensorValue = 0;
    int sensorMapped = 0;

    // Set the serial transfer rate to 31250:
    Serial.begin(31250);

    // Set the analog pin 0 to Input pin:
    pinMode(A0, INPUT);
}

void loop() {
    // Read the voltage on analog pin 0:
    int sensorValue = analogRead(A0);

    // Map the interval 0-1023 to 0-127 (7 bits MIDI word):
    sensorMapped = map(sensorValue, 0, 1023, 0, 127);

    // Use sendCC function to send the MIDI data:
    sendCC(0xB0,0x07,sensorMapped);
}

void sendCC(int cmd, int controlNo, int value) {
    Serial.write(cmd);
    Serial.write(controlNo);
    Serial.write(value);
}

The map function used inside the loop block is important. Since the ADC in Arduino is 10 bits, the resulting digital values will be in range 0 – 1023 for 0 – 5 Volts of voltage. Most MIDI bytes use the resolution of 7 bits, which represents a range of 0 - 127. The map function simply maps the 0 – 1023 range to 0 - 127. In that case, it is possible to say that 10 bits resolution of an Arduino board is exceedingly enough to represent a MIDI byte.

4. Conclusion

The idea of DIY (do-it-yourself), which appeared in 1950s and mostly associated with home improvement and maintenance activities of suburban husband, evolved parallel to today's interests and also became a term referring to any action of designing and building electronic gadgets by someone's own. By means of new generation microcontroller based development boards, it is now possible for many developers to build their own tools as robots, security systems, and control interfaces without the need of any high level expertise and experience. A variety of sensors and these boards, working together, are able to connect our physical environment to the digital world.
The idea of using the information received from the physical world to interact with a digital artwork can be interesting and exciting to an artist. For an electronic musician, human-computer interaction technologies are important tools for translating physical actions into musical expressions. Modern microcontroller based developing boards, when used together with appropriate sensors can be perfect tools to meet the requirements of the contemporary electronic musician.

References


Evans, Martin and Noble, Joshua and Hochenbaum, Jordan. ""Arduino in Action".

Shelter Island NY: Manning, 2013.


Indie games: A case study

KONCA ŞAHER, FERIDUN ÖZIŞ

FACULTY OF ARTS & DESIGN, DEPT. OF INTERIOR ARCH. & ENVIRONMENTAL DESIGN, KADIR HAS UNIVERSITY, İSTANBUL
FACULTY OF FINE ARTS, DEPT. OF MUSIC TECHNOLOGY, DOKUZ EYLÜL UNIVERSITY, İZMİR
konca.saher@khas.edu.tr
feridun.ozis@deu.edu.tr

Abstract

Auralization gives us the opportunity to simulate the acoustical experience at a specific position in the room. It is also a very powerful tool to evaluate the quality of the acoustical environment in a space before it is actually built. It provides a means to detect acoustical defects of a space such as, too much reverberation, focus points, flutter echoes etc. However, the most important question for the architects and the users is: "How do we really hear the space?" There is a growing demand to assess the acoustical parameters subjectively. More clients/architects are demanding auralizations of the rooms from the acoustic consultants during the design stage. The auralizations demonstrate the acoustical quality very strongly. It is a very powerful tool when discussing the acoustical design with the architects. Based on the auralizations architects revise the internal finishes as to create a satisfactory acoustic environment. This paper focuses on how auralizations have affected the design process, especially the choice of internal finishes, for three different types of buildings: First one is an open plan office, second one is a classroom for special educational needs students (SEN) and third one is public address system at underground stations. The auralization samples clearly demonstrated the effect of various internal finishes on room acoustics of these spaces and this had an important impact on the choice of internal finishes by the architects during the design stage.

1. Introduction

Atkins Acoustics, which is part of one of the leading engineering and multi-disciplinary design consultancies in the world with its 17700 employees worldwide, works collaboratively with architects, structural engineers, mechanical engineers, rail engineers and Local Authorities on a wide variety of projects. Room acoustics modelling programs are being used for performance analysis of various rooms such as open plan offices, classrooms, atria, cafeterias, media rooms, performing spaces, recording studios, metro and train stations, airports and etc.
Room acoustical modelling is a process where the behaviour of sound in a room is simulated in a way it would have been in a real space. Acoustical computer simulation programs provide results of several objective acoustical parameters such as reverberation time (RT), sound pressure level (SPL), speech transmission index (STI) for a given three-dimensional model of a room of a certain architectural design. However, the members of the design team, mostly architects, mechanical engineers, interior designers are either not very familiar with these parameters very well or have difficulty in interpreting its effects on quality of sound. Therefore, during the design team meetings/discussions it emerges that there is a necessity for a stronger assessment tool for acoustic quality rather than the objective acoustic quality numbers when communicating with different disciplines. The most important question for the architects and the users is: "How do we really hear the space?"

Auralization gives us the opportunity to simulate the acoustical experience at a specific position in the room. It is also a very powerful tool to evaluate the quality of the acoustical environment in a space before it is actually built. It provides means to detect acoustical defects of a space such as, too much reverberation, focus points, flutter echoes, or delayed sound form coupled volumes. [1].

There is a growing demand to assess the acoustical parameters subjectively. More clients are demanding auralizations of the rooms from the acoustic consultants. The auralizations demonstrate the acoustical quality very strongly. We can investigate the relative differences between various architectural designs during the design stage. This especially affects the choice of internal finishes during the design.

This paper focuses on how how auralizations affected the design process (especially the choice of internal finishes) and the outcoming sound quality, for three different types of buildings: First one is an open plan office, second one is a classroom for special educational needs students (SEN) and third one is public address (PA) system at underground stations. The auralization samples clearly demonstrated the effect of various internal finishes on room acoustics of these spaces and this had an important impact on the choice of internal finishes by the architects during the design stage.

2. Open Plan Office

The client designed an open plan office which is 2600m3 has 20 desks and 80 people working. (Figure 1) The office has quite ordinary internal finishes for an office; painted walls, thin carpet and wooden work stations. However, it does not have any suspended acoustic ceiling which is very common in offices. The office ceiling is an exposed soffit. Design decisions strongly dictate that an exposed soffit (bare concrete structure) is to be used for sustainability reasons (thermal mass heating and cooling). The client was concerned about the background noise levels in the office. As the acoustic consultant of the project, we recommended absorptive panels or integrated systems hung from the ceiling. For absorptive panels "suspended acoustic baffles" or "free hanging acoustic raft" could be a solution. For integrated systems "lighting trays with acoustic elements" could be used. "Suspended acoustic baffles" or "free hanging acoustic raft" options were not suitable for aesthetic reasons. Therefore, "lighting trays with acoustic elements" were recommended in the office to reduce the reverberation time, improve speech intelligibility and in general acoustical quality of the office. The reverberation time (mid frequency) decreased from 1.1 sec to 0.64 sec.
Figure 2 demonstrates the set-up for auralisation in the office. In the set-up we assumed that 25% of the people in the office were talking (10 people are in conversation and 10 people are on the phone). First, auralisation of each person was done at the listener position and then 20 auralisations were mixed to provide the total effect of the background noise at the specific listener position in the office. The same set-up was repeated in the office when "lighting trays with acoustic elements" were applied on the ceiling. Auralisations of the background noise for these two different architectural set-ups were prepared for the client. The auralisations showed clearly how the office will sound differently for the two different architectural set-ups. The client was much more satisfied with the design option with "lighting trays with acoustic elements".

Figure 1 3D model of the open plan office.

Figure 2 Auralisation set-up for two architectural designs; one with the exposed soffit, second one with "lighting trays with acoustic elements".
3. Classroom in a Special Educational Needs School

Our client was in charge of designing a high school for children with special educational needs "SEN". Building Bulletin93 (BB93): Acoustic Design of Schools Design Guide [2] specifies that reverberation time of 0.8 sec (mid frequency range) is the maximum limit for classrooms in high schools in UK. However, when a school is to be designed specifically for children with special educational needs "SEN", we would recommend that the required reverberation time is not to exceed 0.4 sec in mid frequency range to secure a comfortable acoustical environment. Reverberation time of 0.4 sec is the recommended reverberation time for classrooms designed specifically for use by hearing impaired students by Building Bulletin93 (BB93). High percentage of the "SEN" students also suffers from mild hearing impairment. Therefore, reverberation time of 0.4 sec is considered appropriate for "SEN" classrooms. The increase of reverberation time in low frequency should also not increase more than 50% of the mid frequency reverberation time.

Two auralization set-ups were prepared; one with reverberation time of 0.8 sec. (absorptive ceiling) and the second one with reverberation time of 0.4 sec. (absorptive ceiling with high values at low frequencies and absorption on the walls. The auralizations clearly demonstrated the importance of the absorption to reduce reverberation time and absorption in low frequencies in particular. Figure 3 and Figure 4 below shows the 3d view of the classroom and the auralization set-up respectively.

![Figure 3](image1.png) 3D view of the classroom in special educational needs students (SEN) school.

![Figure 4](image2.png) Simulation set-up for the auralization, indicating the listener and source positions.
4. Public Address System of Underground Stations

It is becoming increasingly widespread to use room acoustic modelling programs in design of public address (PA) systems. The public address (PA) system at underground stations in UK is usually required to achieve a minimum speech transmission index of 0.5 at any area of the station. [3]. A system that is both audible and intelligible is required for public address.

The actual reverberation time in all areas of most stations is very high since all the areas are comprised of highly acoustically reflective surfaces such as ceramic tiles and concrete. Therefore, to reduce the reverberation time and increase the speech transmission index in all areas, acoustic treatment should be applied in the form of ceiling absorbers and/or wall absorbers. However, hard surfaces like ceramic tiles, concrete or metal panels are preferred from aestheical point of view as well as ease of cleaning. Materials detailing has to be fit for purpose and robust enough to withstand damage from crowd. Most acoustic treatment would compromise the visual aesthetic.

We recommend acoustic treatment on the ceiling and in the form of perforated panels with acoustic backing behind to most of our clients. Highly efficient acoustic treatment could be achieved with perforated panels backed with acoustic foam, mat etc. This configuration would also not compromise the visual aesthetic very much. However, there is a strong resistance from the architects not to apply any absorption to the station areas. Instead architects mostly argue that adding more speakers, re-arranging the location of speakers or changing the type of speakers should achieve the required speech transmission index in the stations. While some of these arguments might be true at times, in general certain amount of absorption is required to achieve the speech transmission index of 0.5 across any area of the stations. We demonstrate the improvement in quality/intelligibility of speech with application of absorption to our clients by auralizations in many public address system projects. Figure 5 shows an example where auralizations demonstrate the speech quality in a platform when there is no absorptive ceiling and when there is an absorptive ceiling.

Figure 5 Auralization set-up at the underground station platform for two design options with and without acoustic ceiling.
5. Conclusions

In this paper, auralization was discussed as a sound quality assessment tool in practice when discussing the acoustical design with the design team, especially with architects. Based on the auralizations architects revise the internal finishes as to create a satisfactory acoustic environment. This paper focused on how auralizations have effected the design process, especially the choice of internal finishes, for three different types of buildings: First one is an open plan office, second one is a classroom for special educational needs students (SEN) and third one is public address system at underground stations. The auralization samples clearly demonstrated the effect of various internal finishes on room acoustics of these spaces and this had an important impact on the choice of internal finishes by the architects during the design stage.

References


In the present chapter, I wish to unpack a mechanism of listening that is becoming more urgent when dealing with audio technologies in our lives: overhearing, eavesdropping, or in its more technical, pathological expression, ecouterism. I argue that the latter concept (as suggested by Elisabeth Weis in film studies) and related psycho-analytical concepts of listening and overhearing in Critical Theory can help us to understand the aesthetic listening experiences of iPods in audio walks by Janet Cardiff, Dries Verhoeven and Judith Hoffland. With this, I will update Shuhei Hosokawa’s canonical essay, "The Walkman Effect". I will show how a notion of desire of overhearing can explain a specific mechanism in the relation between spectators, "locative" media and the spaces these users dwell in, both physically and mentally, by means of new aesthetic experiences of the "secret theatre".

Abstract

In the present chapter, I wish to unpack a mechanism of listening that is becoming more urgent when dealing with audio technologies in our lives: overhearing, eavesdropping, or in its more technical, pathological expression, ecouterism. I argue that the latter concept (as suggested by Elisabeth Weis in film studies) and related psycho-analytical concepts of listening and overhearing in Critical Theory can help us to understand the aesthetic listening experiences of iPods in audio walks by Janet Cardiff, Dries Verhoeven and Judith Hoffland. With this, I will update Shuhei Hosokawa’s canonical essay, "The Walkman Effect". I will show how a notion of desire of overhearing can explain a specific mechanism in the relation between spectators, "locative" media and the spaces these users dwell in, both physically and mentally, by means of new aesthetic experiences of the "secret theatre".

Introduction

Recently, the wider public has become more aware than ever of the dangers of being overheard through mobile technology, since the whistleblowing of American agencies NSA and GCHQ, the phone tapping in Angela Merkel’s office and more near to home, the leaked wiretaps preceding Turkey’s latest local elections. When combined with locative technology (such as automated geo-annotation through GPS), however, the possibilities for collecting and using metadata (‘data mining’) for security or commercial purposes are unseen, in the double sense of the word. Never in the history of mankind has privacy breach been so much of an ill and an anxiety of the (post-)modern individual who has been entrenched by the neoliberal doctrines of her/his dependence on computer systems, social networks and smart technology.

However, I will not discuss the common beliefs and worries regarding these systems at length. I am rather interested in the artist’s creative use of the technology and the question whether new ‘locative’ audio technologies, such as the iPod and iPod Touch, as explored for aesthetic purposes like audio walks, have indeed new experiences of listening on the offer, particularly with regard to a more self-aware overhearing. This implies a somewhat historical question whether these new locative media only ‘remediate’ the old, semi-private listening experiences, such as the Sony Walkman’s, or if they mediate truly new experiences of image, sound and space relations.
At the core of this discussion lies my interest in the postmodern listening self, which comes to the surface in these hyper-individual relations to these locative media but, equally, to its real and virtual environments the technology helps to mediate and construct, including other listening subjects. The locative aspect of these technologies – with 'locative' hinting at a spatial mode or relation (from the Latin, vestigial case) – does not only concern the where, the place or location, but operates to rebuild and re-establish context of our listening experiences.

Now, to start with the historical question of remediation and the modernist myth of the new, I suggest to look at Shuhei Hosokawa's influential essay, 'The Walkman Effect' (1984, originally written in 1981) not so long after the arrival of the Sony Walkman, in which he describes the influence of the Walkman – or as Michael Bull later proposes, the 'personal stereo' – on the relationship between the individual, 'self-enclosed' listener and urban space. Most notably, in its pragmatic use the Walkman effect turns the device into a secret theatre and its user becomes a secret listener. I intent to take Hosokawa's secret theatre notion further from a similar point of view: "The walkman effect must be measured in terms of this practical mode of operation. Even when one switches off, or leaves it behind, theatrical effects are still active. The show must go on till the death of the gadget-object" (179)."

If we can take this 'death' literally, as the pragmatic functions of the Walkman have been largely overtaken by digital audio media, such as the iPod, we can ask ourselves: how did the show go on, in its 'locative' afterlife? More narrowly, I wish to focus on these media for aesthetic practices, such as audio walks, and their social implications, so I ask myself the question: Which of Hosakawa's statements need revision?

With the rise of a global iPod culture in public, urban spaces, the changing effects of these secret theatres for the ears to our listening modes, spatial relations and ultimately, to ourselves, have been explored by many artists, among which Janet Cardiff has been perhaps the most influential since the end of the 1990s. As a departure point, I take one of her more recent works, Alter Bahnhof Video Walk (????? presented at Documenta 13 in Kassel, 2012), to start exploring the 'secret theatre' concept for its relevance today. After this brief exploration I will continue to discuss two more works, Dries Verhoeven's ???? Niemandsland (2012) and Judith Hofland's ???? Like Me (2013), in which the hyper-individual experience of location is one step further explored in terms of real-time encounters with other people, other 'overhearing' subjects. Most notably, all three works take stations as their sites of departure, from where the audience as individual listening subjects start their sonic journeys.

Introduction

Cardiff and Miller's ???? Alter Bahnhof Video Walk (2012) gives us some first insights into the new 'secret theatre'. The setting for this piece is Kassel's main train station. The participant individually received an iPod, which displayed a visual recording of the itinerary of the sound walk mixed with Cardiff's voice as off-screen narrator. The participant is asked to hold the iPod at all times in such a way to match the given perspective, while the voice guides him the way. By doing so, the iPod turns into an unsophisticated version of locative media. The participant is here the locative agent on command by the artist, turning the logic of 'geo-annotation' in human hands to reveal the exact location of these secret places. The participant becomes a detective following the traces of the hidden mastermind-editor-omniscient narrator personified by Cardiff's disembodied voice.
In my research, I consider ‘acousmatization’ (to denote sound one hears without seeing its source body or originating cause, from Greek *akousma* meaning auditive perception, or literally ‘what is heard’) as essential to understand the impact of sound on listening, since every sound or voice is in a way always acousmatic. Yet, since the beginning of phonographic history, the recording technology has made us only more aware about sound’s ‘eerie’ powers on our cognition and hunger for meaning making. Mobile audio technology has made us more familiar with continuously disembodied sounds while inciting us to re-embody them through our headphones and walk acts; like Hosokawa describes:

"Whether it is the walkman that charges the body, or, inversely, the body that charges the walkman, it is difficult to say. The walkman works not as a prolongation of the body (as with other instruments of musica mobilis) but as a built-in part or, because of its intimacy, as an intrusion-like prosthesis (see Traverses 1979). The walkman holder plays the music and listens to the sound come from his own body (see Barthes 1982, p. 265)” (Hosokawa 176).

Moreover, Hosokawa speaks of a deterritorialized listening as part of the Walkman effect, against R. Murray’s territorialized listening in his attempts to make urban space familiar as a ‘space of security’ (Barthes 1982):

"It intends that every sort of familiar soundscape is transformed by that singular acoustic experience coordinated by the user's own ongoing pedestrian act, which induces an autonomous 'head space' between his Self and his surroundings in order to distance itself from - not familiarise itself with - both of them. The result is a mobility of the Self. Thus the walkman crosses every predetermined line of the acoustic designers. It enables us to move towards an autonomous pluralistically structured awareness of reality, but not towards a self-enclosed refuge or into narcissistic regression” (Hosokawa 175).

Cardiff touches the crux of Hosokawa’s strong belief and deconstruction of the Walkman user’s autonomy, the ‘autonomy-of-the-walking-self’ as urban strategy to contextualise the ever-becoming-complex reality in the 1980s while decontextualizing the city’s given coherence through individual, mobile sound experiences. Hosokawa seemed to respond to the negative outlook on the modern subject, who fails to make coherence. Adorno (2002) refers to the failure in the listener to read music as a meaningful whole. Rather, the modern listener abides by ‘atomistic’ listening. Many scholars after Hosokawa like Michael Bull have described the ‘personal stereo’ as a cinematic experience added like sunglasses to our daily lives and environments, which make them highly personal yet social in terms of establishing new coherence.

Cardiff places now besides the enclosing stereo headphones, a screen between self and surroundings – a screen that has the aura of a memory as much as a vision that sees more. Not only images give perspective on the physical space, also sounds create listening points to the otherwise unseen, offering us this "autonomous pluralistically structured awareness of reality” (Hosokawa). So far, her piece is in line with Hosokawa’s statements.

Through the initial statement “It is very intimate to watch people", the iPod reveals a secret theatre, a promise to a deeper reality that only the bearer of the iPod has access to – it becomes a tool for a secretly shared reality between sounds/voice and self, just as Hosokawa envisaged about the Walkman. Yet this new audio-visually multi-layered and intimate space has also something perverse: watching and overhearing others as they move otherwise unnoticed but who now become part of the participant’s
secret theatre discloses a deeper pathology that lies embedded in our attraction of these new, locative devices: a desire to overhear, oversee and gain secret information to which others have no access, while at the same time being subjected to the technology, as Cardiff remarks, "like those prisoners stuck in Plato's cave".

**Subjectification: Ecouterism and the Art of Overhearing**

As a next step, I will be focusing on two other works from around the same time as Cardiff’s piece: Dries Verhoeven's *Niemandsland* (2012) and Judith Hofland’s *Like Me* (2013), which particularly highlight individual experiences of the self in relation to a new sense of sociability that materializes into real-time urban encounters with places and people. Like Cardiff’s piece, these ‘audio theatre’ pieces turn the privacy of the highly-individual experience of the secret theatre into a feeling of submission (a feeling of being subjected) to a piece of technology and a disembodied voice. Yet both focus on different experiences of human contact through breaking the distance.

Dries Verhoeven’s *Niemandsland* (2012) shows us how distance can energize a socially engaged experience in a common yet secretly shared space. In *Niemandsland* (literally, "no man’s land"), each participant is paired like in a speed date or blind date to a stranger, a foreigner, someone with a migration background, a history of diaspora and displacement. They meet each other, while wearing headphones and an iPod, in a station (in the Netherlands, it was originally done in Utrecht and Amsterdam Sloterdijk). Their random encounter, invisible to passers-by, turns into a pathway through a multicultural neighbourhood as much as a highly personal testimony in the lives of the other. The iPod is used here to show a slice of reality that goes otherwise unnoticed. Responses by participants range from a feeling of recognition to powerlessness against the social injustices they hear about. Dispossession here means literally the removal and exclusion from urban space as we know or assume to know it.

This audio walk or ‘theatre’ turns the sense of being ‘in charge’ as iPod user over one’s environment around and plays on a voyeurism in the mobile listening act, enhanced by the secrecy of the headphones which make one feel safe to look shamelessly at others who become part of one’s own secret theatre. This is what Michael Bull calls ‘auditized looking’, namely the ability of those listening to a personal stereo to make or escape eye contact with others in ways they would not otherwise.

Elisabeth Weis refers to this privileged listener as the écouteur (the eavesdropper, from French écouter, listening) as equivalent to the voyeur. Hence, she coins écouterism to describe the pleasure in aural stimulation as equivalent to voyeurism. To her, this phenomenon is central to the cinematic experience. Since mobile personal stereo devices such as Walkman and iPod have been compared to a cinema of the mind, I suggest to extend Weis’ notion – and thereby Hosokawa’s analysis – to the secret listening situation. She states: "In every case the eavesdropper acquires some form of knowledge... a self-knowledge that the listener would not otherwise have recognized" (Weis n.p.).

She further refers to the so-called "primal scene" when Freud describes the impact of the child overhearing his/her parents engaging in sexual intercourse. She adds "It is often not important what words are overheard; rather, that knowledge is often of something momentous, terrible (anxiety producing), erotic, and secret – carnal knowledge" (Weis). It is as such that the iPod in *Niemandsland* produces very intimate,
embodied experiences, which are playful, like a dance in urban space, and at the same time 'terrible' when overhearing someone's life story and watching him/her from behind.

The distance is kept, even the voice is admitted to be one of an actor, which functions like a mask or a wall, and enhances the voyeurism, until the invisible curtain is dropped: the other person turns around and looks straight at you of which there is no escape. This moment reminds me of the description in a passage of Being and Nothingness, where Sartre recounts how a man peeps through a keyhole, which completely absorbs him in his investigating gaze towards what is unfolding behind the door. Suddenly, the man hears footsteps in the hall behind him. The resounding though invisible footsteps are threatening to the voyeur’s gaze, as Dennis Hollier (2004) comments:

The gaze of the other, as Lacan praised Sartre for emphasizing, has entered the voyeur’s field of non-vision: it is an (offscreen) acoustic gaze, one experienced not visually, but acoustically, through the surprise of hearing another presence, of feeling him there acoustically, through one’s ears. (164)

Sound can arrest your presence, pin you to a location, when it breaks through this fourth wall of secret listening. In the case of Niemandsland, it is the voyeuristic gaze turned round, which shocks the secret listener. When noticed, this precisely makes up the thrill of the secrecy of the Walkman effect.

Breaking the Fifth Wall: Auditory Gaze or Navel-Gazing?

What Dennis Hollier calls the lure and the shock of the auditory gaze as opposed to Bull’s auditized looking is primal to the way we sense ourselves as listening subjects, or as Steven Connor calls it: the modern auditory 'I' or self as way of embodiment and subjectivity in relation to the sounding world. Connor refers to Don Ihde:

‘My "self"’, declares Don Ihde, the most enthusiastic of audiophile philosophers, 'is a correlate of the World, and its way of being-in that World is a way filled with voice and language. Moreover, this being in the midst of Word [sic!] is such that it permeates the most hidden recesses of my self.' (Ihde qtd. in Connor 1997: 219).

Through the mediation of the headphones and technology, audio walks can give exactly this awareness of the auditory self, between the inside and outside world, full of sound and voice.

Judith Hofland’s piece, Like Me (2013) takes this being in the midst of World/Word further by relating it to the voices we imagine when reading words in social media. This interactive audio tour – done for Festival aan het IJ in Amsterdam around the Central station – also makes use of the iPod – in this case, an iPod Touch – which makes it possible to find and meet other spectators equipped with the same equipment. Like Me – in the double sense of 'liking' – makes us think how virtual life and the constant data stream of social media permeate our lives, changing our real encounters and interactions with people.

Here too the use of the iPod plays with the tension between surrender and control, which Hosokawa referred to as part of his autonomy concept. French sociologist Antoine Hennion referred to it as the iPod listener’s own construction of his passivity through his own control:
So instead of popular music pacifying the listener, as Adorno describes it, the listener actively chooses to surrender to the music and create a personal and emotional experience. As such the listener is very much in control throughout the listening. Even when he himself might describe it as 'losing himself in the music' the experience will often be staged and designed. ((belirtilmeyen kaynağa atıf)Gomart & Hennion: *Music Lovers*, qtd. in Juslin & Sloboda, *Music and Emotion*; in Leong & Gram n.p.)

In Judith Hofland’s piece, the participant wilfully surrenders to a disembodied voice, called 'Sascha' who feeds him with options, possible perspectives and philosophical thoughts on his pathway through the station, while becoming a member of the secret voice’s social network, activating him. While walking, the participants answer questions about themselves on the touch screen and make connection with invisible, online friends, whose presence is marked through messages. Then, Sascha introduces him to one other person in the network through a picture one chooses from three options, taken from the web. They are both asked to give their first impression on the basis of the photo. Then suddenly, they meet physically.

Whereas *Niemandsland* broke with the proverbial fourth wall in making the audience aware of its own perception reaching through the invisible canvas that separates life from theatre, *Like Me* goes one step further by breaking the fifth wall, namely the metaphorical barrier in our interaction with the media and ourselves or "that semi-porous membrane that stands between individual audience members during a shared experience" ((references’da yok)Davenport 2000 et al.). Through these pieces, we make the journey from an inwardly-drawn flâneur or secretive écouteur, dwelling in the colours or moods of their environments as a shared secret between them and the artist, to a socially charged environment, where the participant is given social agency and becomes an active ‘spect-actor’ (Boal) or perhaps ‘aud-actor’ (my term). In the end, the final question is being asked, 'who do we prefer: the virtual Sascha or the real person?'

The digital afterlife of the Walkman in the iPod begs the question: do we want still the anonymity, the privilege and appeal of the secrecy? According to Hosokawa a comparable question had already become obsolete by the time of his article on the "Walkman Effect" in 1984, in which he reacts against all cultural pessimism that the Walkman would induce asocial behaviour, rather:

*Autonomy is not always synonymous with isolation, individualisation, separation from reality; rather, in apparent paradox, it is indispensable for the process of self-unification. Walkman users are not necessarily detached ('alienated' to use a value-laden term) from the environment, closing their ears, but are unified in the autonomous and singular moment - neither as persons nor as individuals - with the real. (Hosokawa 170)*

Let me repeat with Hosokawa: "It enables us to move towards an autonomous pluralistically structured awareness of reality, but not towards a self-enclosed refuge or into narcissistic regression" (175). And yet, today artists pose again the question with regard to social media and the way we make use of them through our iPhones and iPods: aren’t we all engaged in narcissistic navel-gazing under the pretext of being social?
Judith Hofland’s interactive piece makes us see that these questions about how we relate to each other through social networks are intrinsically connected with how we occupy public space, how we ’sound out’ our existence and welcome others in our private spaces. And this is precisely what locative media are about, according to Drew Hemment (2005): ”Locative Media foregrounds place and context within social and creative interventions, and explores ways in which the devices people carry, the fabric of the urban environment and even the contours of the Earth may become a digital canvas.”

Hosokawa’s deconstructive but positive stance towards the Walkman can give us some hope on how this digital canvas can be used as a creative strategy to re-socialize and reconnect with a new construction of collectivity and sociability. As locative media are becoming part of our everyday lives and the controversies of our social existence, Hofland gives us things to think about. But Hosokawa showed us, once more, how intimate relational experiences through the iPod can open our ears and eyes to our socially complex reality, rather than to shut them off.

References


Audio Feature Extraction for Exploring Turkish Makam Music

HASAN SERCAN ATLI, BURAK UYAR, SERTAN ŞENTÜRK, BARİŞ BOZKURT, XAVIER SERRA

“BAHÇEŞEHİR UNIVERSITY, UNIVERSITAT POMPEU FABRA-BARCELONA, KOÇ UNIVERSITY

hsercanatli@gmail.com, burakuyar@gmail.com, sertan.senturk@upf.edu, barisbozkurt0@gmail.com, xavier.serra@upf.edu

Abstract

For Turkish makam music, there exist several analysis tools which generally use only the audio as the input to extract the features of the audio. This study aims at extending such approach by using additional features such as scores, editorial metadata and the knowledge about the music. In this paper, the existing algorithms for similar research, the improvements we apply to the existing audio feature extraction tools and some potential topics for audio feature extraction of Turkish makam music are explained. For the improvements, we make use of the Turkish makam music corpus and the culture specific knowledge. We also present a web-based platform, Dunya, where the output of our research, such as pitch histograms, melodic progressions and segmentation information will be used to explore a collection of audio recordings of Turkish makam music.

1. Introduction

To gain an analytical understanding about a music tradition and the relations between its descriptive attributes, it is useful and practical to get the help of the computational methods. Taking the advantage of the improvements in information retrieval and signal processing areas, we can organize large amount of data, navigate through large collections and discover the music genres, etc. By using the computational methods, more data can be processed in less time compared with the manual research methods. What is more, music exploration systems can be designed specifically for different traditions according to their specificities.

A definition for an exploration system is "a platform that provides new technologies, interfaces and navigation methods to browse through collections from specific music cultures" (Porter et al 101). A music exploration system allows users to reach the musical content in a structured way or provides tools to analyze the attributes of the music. This structure can be designed according to the aim of the system and can be organized in many different approaches. Currently, there are several examples for the music exploration systems. One of them is Sonic Visualiser (Cannam et al 324), which is an exploration system that visualise low-level and mid-level features from the uploaded audio recording. Another music exploration system is Musicsun (Pampalk and Goto 101), designed for artist recommendation.
Our study is the Turkish makam music branch of the CompMusic project (Serra 151). In the CompMusic project, our aim is to explore music cultures other than the Western popular by taking advantage of the advancement in Music Information Retrieval. We work on extracting musically meaningful descriptions from audio recordings by using research tools and support these descriptions with additional information, such as music scores or editorial metadata. Besides this, we develop an application Dunya (Porter et al 101), where we can evaluate our research results from a user perspective. Dunya is a web-based exploration system that users can reach our research results easily.

For this kind of research and studies, there is a need for a corpus which should mainly include audio recordings. By analyzing the audio recordings using computational methods, different aspects of a music tradition can be understood. In addition to the audio recordings, it is beneficial to have related supportive information from the music tradition (e.g. culture-specific knowledge). This information helps studies to be multi-directional and understand the connections between them. At the same time, the corpus should contain data as much as possible with a high degree of diversity in order to represent the tradition well enough for research. For these reasons, we use Turkish makam music corpus (Uyar et al 57) in our research. This corpus includes audio recordings, music scores and editorial metadata. By using the different data types in this corpus we are able to conduct our experiments based on the features we are interested in.

In the context of music, features give semantic information of the analysed material. In our methodologies, we use features and other relevant information (e.g. metadata, music scores and culture-specific knowledge) to understand the characteristics of Turkish makam music. Because the features are our main focus in our study, we have created a feature list to decide and classify the attributes we plan to extract from the corpus to discover Turkish makam music tradition. By using this list, we can prioritize and follow-up our experiments in a structured way. While developing our methodologies, we examine the existing studies on similar problems and, if available, we modify the best-practices aptly to our case. In this paper, we present the extracted features and the features we plan to extract for Turkish makam music to include in Dunya. Moreover, we explain new methodologies for extracting the predominant melody and pitch distribution.

The paper is structured as follows: In Section 2, we briefly explain the key attributes and characteristics of Turkish makam music. In Section 3, we describe the corpus we are using for our research. In Section 4, our music exploration platform, Dunya, is presented. In Section 5, we explain the features related to our study and in Section 5, finalize the paper with a brief conclusion.

2. Turkish makam music

In Turkish makam music, there are three main concepts to explain the main attributes of the pieces. These concepts are makam, usul and form. Makams mainly constitute the melodic aspects and are modal structures which have initial, dominant and final tones. The melodic progression starts around the initial tone, moves around to the dominant and goes to the final tone in the end, which is also called as tonic. These tones and the melodic progression are used to describe a certain makam. Usuls include strokes with different velocities to describe the rhythmic structure. Usuls can be very basic with one strong and one weak stroke or a specific usul can include the combination of multiple usuls.
In the practice of Turkish makam music, musicians have large room for interpreting a music score and may add many ornaments while playing. With this concept, Turkish makam music is more of an oral tradition. On the other hand, this tradition has been represented by using a modified Western notation for the last century. The scores include the main melodies and musicians interpret the scores with their own approaches. The most accepted theory is Arel-Ezgi-Uzdilek (AEU) (Arel). This theory approximates to the 24 tones in an octave by using the 53-TET system, which divides a whole tone into 9 equal pieces.

Another characteristic of Turkish makam music is heterophony. In the performances the musicians play the melody in the register of their instruments or vocal range. However, each musician applies their own interpretation to the melody by adding embellishments, expressive timings and various intonations. The musicians might adjust tuning of the instruments among each other according to the makam, or personal taste. The tonic frequency is also not definite as it may be adjusted to a number of different transpositions, any of which could be favored over others due to instrument/vocal range or aesthetic concerns.

3. Turkish makam music corpus

Within the scope of the CompMusic project, one of the most important tasks is to create music corpora, which represent the music traditions to be studied. The aim of preparing such corpora is to facilitate research on the music traditions by providing well-structured and representative data. These corpora are tailored considering the criteria of purpose, coverage, completeness, quality and reusability (Serra 1).

In the Turkish makam music corpus (Uyar et al 57), there are three main types of data: audio recordings, music scores and editorial metadata. There are 5953 audio recordings in the corpus, which consist of commercial/non-commercial releases as well as bootleg concert recordings. 150 distinct makams, 88 usuls and 120 forms are performed in the audio recordings. In the score corpus, there are 2200 machine-readable score files from 157 makams, 82 usuls and 64 forms. The main source for the metadata is MusicBrainz and there are ~27000 entries related to the corpus. These entries include all available information about the entities such as album cover information, biographies, lyricists and makams. These relationships allow the researchers to use different types of data sources or combine them for a specific study or experiment. Some of the possible studies are explained in Section 5.
To easily access the data sources and the relationships, we make use of a simple database which stores the audio recording path, related score file's path, MBID of the recording, MBID of the related work on MusicBrainz and the related culture-specific metadata, (i.e. *makam*, *usul*, *form*). By using MBIDs, the metadata on the album covers and the detailed information of the metadata can be accessed as well, (*e.g.* biography of the artist).

4. Dunya

Dunya is planned to be a platform where the outcomes of the research in scope of the CompMusic project are going to be presented. By using the data provided on Dunya, a user who wants to learn Indian or Turkish makam music tradition briefly can reach to the basic information or a more experienced user can use the research results to explore the analytic structure of this music tradition. Mainly, the data provided on Dunya represents the *melodic*, *rhythmic* and *structural* properties of the records. With this information, a user can improve his/her understanding of that certain music tradition in an analytical approach.

![Figure 2 Dunya-Makam mock-up](image)

For Turkish makam music version of Dunya, a mock-up for the recording page is provided in Figure 2. Recording page is the main interface for this tradition because recordings are the main entities in Turkish makam music. On a recording page, the melodic, rhythmic and structural features of the recording are presented. In this context, the interface can be examined under three parts. In the topmost part, pitch distribution, predominant melody and the tonic of a certain recording are presented, which are related to the melodic attributes. For the rhythmic attributes, the beat/downbeat information with usul cycles is displayed on the topmost part. In addition to the rhythmic analysis of the recording, the related *usul* is presented in the lowest part,

---

1 MusicBrainz Identifier
where user can listen and see the strokes. In the middle, there are the score and lyrics of the piece, which are computationally aligned with the audio and include the section and phrase information. Section and phrase information are related with the structural attributes of the audio. These features of a recording are helpful to understand a piece and the related makam and usul as well.

In our research, we compute and analyze the relevant features of the elements in our corpus to understand the tradition. This helps us to discover both high and low level attributes of the music tradition and the relations between different elements in it (e.g. *makams, usuls*). A similar design for Carnatic and Hindustani music traditions has already been implemented and can be seen on Dunya[^2].

### 5. Features

In the CompMusic project, we mainly focus on the extraction of melodic, rhythmic and structural aspects of the studied music traditions (Serra 151). For Turkish makam music, we use the audio recordings, music scores, related metadata available in the corpus and the knowledge provided by the masters of this music tradition.

In Table 1, we present a list of features for Turkish makam music. This list consists of the features that we have already extracted by running relevant algorithms on the Turkish makam music corpus and those we aim to reach within the scope of the CompMusic project. They have been classified under three categories as melodic, rhythmic and structural. In the rest of this section we explain these categories in detail.

#### 5.1. Melodic Features

In our studies we extract the features such as predominant melody, pitch distribution, tonic and makam. Using these features, we can analyze the melodic progressions, the similarity between makams or the style of a certain performer or composer etc.

##### 5.1.1. Predominant Melody

In the analysis of eurogenetic musics, chroma features are typically used due to their ability to represent harmonic content and their robustness to noise and changes in timbre, dynamics and octave-errors (Gómez). On the other hand, predominant melody is preferred to study the melodic characteristics of Turkish makam music due to its heterophonic nature (Bozkurt et al 3). (Gedik and Bozkurt 1049) uses YIN (De Cheveign et al 1917) to estimate the fundamental pitch and then post-processing step to correct octave errors and short erroneous jumps. While YIN outputs accurate pitch estimations for monophonic recordings, in (Şentürk et al 57) it is observed that it does not output reliable estimations for heterophonic recordings.

[^2]: [http://dunya.compmusic.upf.edu/carnatic](http://dunya.compmusic.upf.edu/carnatic)
(Şentürk et al 34) uses the methodology proposed by (Salamon and Gomez 1759) to extract predominant melody. Figure 3 shows the steps followed to compute the predominant melody. Note that the methodology proposed by (Salamon and Gomez 1759) is optimized for popular musics with a predominant melody and accompaniment such as Western pop and jazz. The methodology assumes that there is no predominant melody in time intervals where the peaks of the pitch saliences are below a certain magnitude with respect to the mean of all the peaks. Moreover, it eliminates pitch contours, which are considered as belonging to the accompaniment.

Since time intervals without predominant melody are rare in Turkish makam music, the methodology with default parameters erroneously discards a substantial number of pitch contours in Turkish makam performances. (Şentürk et al 34) changes some parameters according to the specificities of Turkish makam music to overcome this problem. In the structure-level audio-score alignment experiments, predominant melody computed with modified parameters yield better results compared to YIN and chroma features.

On the other hand the predominant melody computed with modified parameters still produces substantial amount of errors when the music is played softer than the rest of the piece. This becomes a noticeable problem in the end of the melodic phrases, where musicians choose to play softer. For this reason we decided to optimize the methodology of (Salamon and Gómez 1759) step by step. We first estimate the pitch contours and then use a simpler pitch contour selection methodology, which does not consider accompaniment, to obtain the predominant melody. We utilize Essentia (Bogdanov et al 493) to compute pitch contours. The implementations of this methodology and the one used in (Şentürk et al 34) are available in pycompmusic library.

\[\text{Figure 3} \quad \text{Flow diagram of predominant melody computation using the methodology proposed by (Salamon and Gomez 1759). The names in blocks refer to the corresponding functions in Essentia.}\]

\[\text{Figure 4} \quad \text{Contour bins and predominant melody}\]

\[\text{\url{http://essentia.upf.edu/documentation/reference/std_PitchContours.html}}\]

\[\text{\url{https://github.com/MTG/pycompmusic}}\]
In the computation of the pitch salience function\(^5\) we select the bin resolution as 7.5 cents instead of 10 cents. 7.5 cents approximately corresponds to the smallest noticeable change ($\frac{1}{3} Hc$) in makam music (Bozkurt 13). In the computation of pitch salience peaks, we experimented on different values of the peak distribution threshold parameter to get a satisfactory pitch contour length. Currently, we lack the ground truth to empirically find an optimal value for this parameter. Hence, we decided to set the `peakDistributionThreshold` parameter as 1.4, instead of using the default parameter "0.9" to get and observe longer pitch contours (Figure 4).

Once the pitch contours are obtained, we order the pitch contours according to their length and start with selecting the longest one. Then, we remove all portions of pitch contours which overlap with the selected pitch contour (Figure 4). We carry the same process for the next longest pitch contour, and so forth. By repeating the process for all pitch contours, we obtain the predominant melody of the audio recording (Figure 4). Some predominant melodies might contain octave errors because of the heterophony of Turkish makam music. In the future we will implement the postprocessing step used in (Gedik and Bozkurt 1049)\(^6\).

Apart from the predominant melody estimation in audio recordings, we also extract a synthetic predominant melody from the notes and their durations given in the music scores. This feature may be synthesized either according to the theoretical pitches (e.g. according to AEU theory) or according to a tuning obtained from one or multiple audio recordings (Şentürk et al 95) (Bozkurt 43). This feature is currently used in audio-score alignment (e.g. Şentürk et al 43 and Şentürk et al 57). We will also use this feature to playback the music scores in Dunya.

### 5.1.2. Pitch and Pitch Class Distributions

Pitch distributions and the "octave-wrapped" pitch-class distributions are the features commonly used for tuning analysis (Bozkurt et al 45), tonic identification and makam recognition (Gedik and Bozkurt 1049). These distributions typically have a high pitch resolution to capture the tuning characteristics specific to the Turkish makam music. The pitch distributions are useful to capture the tuning characteristics of a recording or a set of recordings (e.g. in the same makam) spanning to several octaves, whereas pitch-class distributions are more desirable for tasks which would suffer from octave errors. These distributions are typically computed from predominant melody. There are two common methods to count the number of samples in the predominant melody that fall into each bin, (histogram) (Gedik and Bozkurt 1049) and kernel-density estimation (Chordia and Şentürk 82).

For each audio recording we use the predominant melody explained in Section 5.1.1 to compute the four possible combinations; namely pitch histogram, pitch-class histogram, pitch kernel-density estimate and pitch-class kernel-density estimate. The bin size is kept as the same as the pitch resolution of the predominant melody (7.5 cents) resulting in a resolution of 160 bins per octave.

We use the intonation library \(^7\) to compute the kernel-density estimates. We select a normal distribution with a standard deviation (kernel width) of 15 cents as the kernel.

---
\(^5\)http://essentia.upf.edu/documentation/reference/std_PitchSalienceFunction.html
\(^6\)http://essentia.upf.edu/documentation/reference/std_PitchFilterMakam.html
\(^7\)https://github.com/gopalkoduri/intonation
In (Şentürk et al 175), this value was empirically found to be optimal for the score-informed tonic identification task. Next, we will select the appropriate distribution and optimize the parameters for other computational tasks and also for the visualizations in Dunya. The code to extract the pitch and the pitch-class distributions are also available in pycompmusic\(^8\) library.

We also use the peaks observed in the pitch (and pitch-class) distributions to obtain the tuning. We use “slope” method in the "GetPeaks" algorithm which is included in pypeaks library\(^9\). We will use this information for adaptive tuning as explained in (Bozkurt 43).

\[ \text{Audio Recording} \]

\textbf{Figure 5} Flow diagram of pitch histogram calculations

\subsection*{5.1.3. Tonic and Makam}

Bozkurt finds the tonic frequency by estimating the frequency of the last note in a recording (Bozkurt 1). Nevertheless, this method is prone to the quality of the predominant melody and impractical for the recordings which do not end in the tonic frequency (e.g. recordings ending with fade-out or applause).

Another approach is comparing the pitch distribution computed from an audio recording and the template pitch distribution of several makams (Gedik and Bozkurt 1049). The audio pitch distribution is pitch-shifted and a similarity score is computed according to each template distribution for each of these shifts. Assuming that the highest score would be observed between the audio pitch distribution and the template of the "true" makam when the shifted tonic and the tonic in the template are matched, the tonic frequency and the makam of the audio recording are jointly estimated.

When a machine-readable score of the performed composition in an audio recording is available, the symbolic melody information can be used to assist the tonic identification task. (Şentürk 175) extract a predominant melody from the audio recording and compute a pitch-class kernel-density estimate. Then the peaks of the estimate are picked as possible tonic candidates. Assuming each candidate is the tonic, the predominant melody is normalized such that the candidate tonic is assigned to zero cents. Then for each candidate the score and the audio recording are partially aligned to each other. The tonic is estimated as the candidate which yields the most "confident" alignment. This method outperforms (Gedik and Bozkurt 1049). However, this method cannot be applied to all audio recordings since it requires the score of the performance to be available.

\(^{8}\)https://github.com/MTG/pycompmusic

\(^{9}\)https://github.com/gopalkoduri/pypeaks
5.2. Rhythmic Features

(Srinivasamurthy et al 94) have been working on rhythmic feature extraction, including Turkish makam music and Indian art music. In their study, they define 3 rhythm related tasks, beat tracking, meter estimation and beat/downbeat detection. For Turkish makam music, currently we are planning to include beat/downbeat analysis for the audio recordings. By using the output of this analysis, the usul structures and the relations between different usuls can be understood deeper.

5.3. Structural Features

In the score corpus, section information of 2200 compositions is available in our machine-readable file formats (e.g. 1.Hane, Teslim). Additionally, for the compositions with vocals, each phrase of the vocal line is marked by space character in SymbTr. Moreover, in (Karaosmanoğlu et al 10), 899 of the scores from this collection have been segmented into approximately 30000 phrases by three Turkish makam music experts. These annotations have been used for (Bozkurt et al 1) study.

(Şentürk et al 57) uses the section information given in the SymbTr-scores to link each section to the time-intervals they are performed in the audio recordings. Next, each section and the time intervals are aligned in the note-level (Şentürk 57). We plan to use the alignment results to estimate the average and local tempo, to automatically retrieve the performances of a selected part from a composition and also to study the tuning and intonation characteristics. By using the lyrics information in the SymbTr-scores, lyrics-to-audio alignment is also studied by (Dzambazov 61).

6. Conclusion

In this paper we have presented a music exploration system for Turkish makam music, Dunya, and the feature extraction methodologies for this music tradition. We have provided the musically descriptive features, which will be computationally extracted from Turkish makam music corpus and will be presented in Dunya. We have presented a mock-up for Dunya-Turkish makam music, explaining how it includes the extracted features. For the existing features, we have provided brief information and references. As being the new methodologies we implemented, flow diagrams and calculations are explained for the predominant melody extraction and the pitch histogram calculation in Section 5.1.1. and Section 5.1.2., respectively. We also provide a list of the musically descriptive features of Turkish makam music. We expect these studies facilitates academic research in several fields such as music information retrieval and computational musicology.

7. Acknowledgements

This work is partly supported by the European Research Council under the European Union’s Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

---

10 https://github.com/MTG/SymbTr
tion of pitch contours from an ordering by the methodology of Juan and Gómez (1999) for infant melody estimation by the pro-posed methodology in Section 4.1.1. This Predominant Melody with the methodology of Sentürk et al. (2007) was found to be predominant melody with the methodology in Sec.
References


Arel, H. S. "Türk Musikisi Nazariyat (The Theory of Turkish Music)". İTMKD Yayınları, 1968


