Technical talks

1. Do Research
2. Write
3. Paper
4. Present
5. Talk
Key Idea

Do Research -> Tell a Story -> Paper

Re-telling Same Story

Talk

Different versions
Same Story, Different Retelling

Main ideas

Motivation

High-level

Talk

Paper

Algorithms

Experiments

$\int \mathbf{B} \cdot d\mathbf{A} = 0$

$\int \mathbf{E} \cdot ds = -\frac{d\Phi_B}{dt}$
Talk: Storytelling
Choose a path from root to leaf

Omit branches
MOTIVATION

Some people need more than others...
MATH: LESS IS MORE

Abbrev: only if well-known
Avoid: SSFX vs FSXF
Additional slides

- Hide them at the end
- Contain additional proofs, experiments, diagrams, charts, etc.
- Anticipate questions
Outline

- Introduction
- Problem Statement
- Literature Review
- Our Method
- Experiments
- Conclusion
Road Map

Content --> Audience --> Medium --> Example
Put humans in a dimly lit, cozy room, with a constant background drone

What happens?
Human Psychology

Limited short-term memory
• Remembers 7 ± 2 things

Short attention span
• “Tunes out” quickly if nothing interesting

Visual-Aural receptiveness
• Responds to Visual + Aural stimuli
• Responds to eye contact
5 ways to put audience to sleep

- Speak inaudibly: mumble
- Maintain monotonous voice
- Fill slides with lots of equations and text
- Avoid eye contact
- Hide behind rostrum
5 ways to engage audience

Dress smartly and conservatively
5 ways to engage audience

Speak clearly

- pronounce words
- project voice
- vary pace & pitch
5 ways to engage audience

Avoid visual overload

*minimize symbols*

*use images, icons instead*
5 ways to engage audience

Look at audience: left, back of room, right
5 ways to engage audience

Move around, smile

*but not too much!*
Repetition

Introduction

Body

Summary

Tell them what you’re going to tell them

Tell them

Tell them what you told them
Handling Q & A

No questions?

*Usually means a boring talk*
QUICK TO LISTEN
SLOW TO SPEAK
SLOW TO ANGER
Handling Q & A

Repeat/rephrase question

Clarifies the question
Allows others to hear it
Buys you time
Don’t overrun your allotted time
Be flexible to adjust your pace
Don’t let difficult questions derail your talk
Road Map

- Content
- Audience
- Medium
- Example
Paper vs. Talk

**Paper**

- Offline, passive
- No speaker; no sound
- Cross-reference possible
- Paper is *dead tree*, not interactive

**Talk**

- Real-time
- *Interactive* Speaker
- Limited X-ref
- Sound + animation
Bullets kill!
Fonts

Arial, Verdana
Arial, Verdana
Arial, Verdana

Times Roman
Times Roman
Times Roman
Colors

- Dark background, white words, OR
- White background, black words

- Avoid gaudy colors
We rendered each face under varying illumination and pose.

Illumination: single light source placed from left to right at increments of 20°, and from bottom to top at increments of 20°

Pose: camera placed from left to right at increments of 20°, and from bottom to top at increments of 20°
Animation + Video
Example

Music Transcription Using an Instrument Model

Jun Yin, Terence Sim, Ye Wang and Arun Shenoy
ICASSP 2005
Music Transcription

Audio signal

Music score

Synthesis

Easy!

Transcription

Hard!
Alternative notation

- MIDI format
  - Musical Instrument
  - Digital Interface
- Well-established “encoding”

<table>
<thead>
<tr>
<th>Onset</th>
<th>Duration</th>
<th>Pitch</th>
<th>Loudness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29</td>
<td>20</td>
<td>1.5278</td>
</tr>
<tr>
<td>26</td>
<td>30</td>
<td>22</td>
<td>1.4738</td>
</tr>
<tr>
<td>52</td>
<td>30</td>
<td>20</td>
<td>1.4726</td>
</tr>
<tr>
<td>52</td>
<td>30</td>
<td>24</td>
<td>1.4952</td>
</tr>
<tr>
<td>77</td>
<td>31</td>
<td>22</td>
<td>1.4188</td>
</tr>
<tr>
<td>77</td>
<td>31</td>
<td>25</td>
<td>1.4322</td>
</tr>
<tr>
<td>103</td>
<td>30</td>
<td>27</td>
<td>1.4605</td>
</tr>
<tr>
<td>129</td>
<td>30</td>
<td>29</td>
<td>1.4593</td>
</tr>
</tbody>
</table>
Basic music terminology

Musical Scale
- A3=220 Hz
- Exponentially Stepped
- Semitone Step= $\sqrt[12]{2}$
- Octave Step= 2

<table>
<thead>
<tr>
<th>Note</th>
<th>Freq (hz)</th>
<th>Note</th>
<th>Freq (hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3</td>
<td>$A3 \times 2^{(0/12)}=220$</td>
<td>C#4</td>
<td>$A3 \times 2^{(4/12)}=277$</td>
</tr>
<tr>
<td>A#3</td>
<td>$A3 \times 2^{(1/12)}=233$</td>
<td>D4</td>
<td>$A3 \times 2^{(5/12)}=294$</td>
</tr>
<tr>
<td>B3</td>
<td>$A3 \times 2^{(2/12)}=247$</td>
<td>D#4</td>
<td>$A3 \times 2^{(6/12)}=311$</td>
</tr>
<tr>
<td>C4</td>
<td>$A3 \times 2^{(3/12)}=262$</td>
<td>E4</td>
<td>$A3 \times 2^{(7/12)}=330$</td>
</tr>
</tbody>
</table>
Basic music terminology

- **Musical Sound**
  - Series of Sinusoid Waves
  - Fundamental = F
    - Related to pitch
  - Harmonics = kF, k integer
  - Harmonic Structure: characterizes an instrument

<table>
<thead>
<tr>
<th>Freq</th>
<th>Amp</th>
</tr>
</thead>
<tbody>
<tr>
<td>220</td>
<td>50</td>
</tr>
<tr>
<td>440</td>
<td>20</td>
</tr>
<tr>
<td>660</td>
<td>50</td>
</tr>
<tr>
<td>880</td>
<td>10</td>
</tr>
</tbody>
</table>

Harmonic Structure: [1, 0.4, 1, 0.2]
Basic music terminology

- **Monophonic**: 1 note at a time
  - No simultaneous notes
  - Transcribing this is relatively easy

- **Polyphonic**: many notes together
  - Harmonic structure overlap!
  - e.g. A3 + A4
    - $(220, 440, 660, 880, \ldots) + (440, 880, \ldots)$
  - e.g. C4 + E4 (some harmonics are close together)
    - Hard to decipher
Key Idea

- Use model of instrument to disambiguate

- Assume harmonic structure
  - Constant across pitch
  - Constant over time
  - Only 1 sample required
  - True for certain instruments, e.g. piano

- Search for harmonic structure in audio signal
Method

1. Create *frequency spectrum* from input audio and instrument sample
Method

2. Create *musical spectrum* from frequency spectrum
   Discretize to 1496 bins
   (88 pitches * 17 harmonics)

3. Match using *spectrum subtraction algorithm*
   -- estimates *pitch* and *loudness*
Spectrum Subtraction Algorithm

Input $Z_M$

$\text{Ins. model } I$

Slide Match Output

(a=1, p=37)

(a=0.8, p=40)
System Implementation

4. Detect *onset* and *duration*

5. Output table

<table>
<thead>
<tr>
<th>Onset</th>
<th>Duration</th>
<th>Pitch</th>
<th>Loudness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29</td>
<td>20</td>
<td>1.5278</td>
</tr>
<tr>
<td>26</td>
<td>30</td>
<td>22</td>
<td>1.4738</td>
</tr>
<tr>
<td>52</td>
<td>30</td>
<td>20</td>
<td>1.4726</td>
</tr>
<tr>
<td>52</td>
<td>30</td>
<td>24</td>
<td>1.4952</td>
</tr>
</tbody>
</table>

6. Convert to MIDI file
## Some Results

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Segment 1</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Minuet in G Major</strong></td>
<td></td>
</tr>
</tbody>
</table>
System performance

- Overall Precision: 0.96
- Overall Recall: 0.98

Performance not affected by
- The duration of the note
- The number of simultaneous notes
- The instrument of the music, as long as the correct instrument model is used

Performance degraded by
- The pitch of the note is too low
- The instrument harmonic structure differs from that in the music
Main Contributions

- Proposed to use Instrument Model for transcription.
  - Disambiguates between overlapping harmonics
  - Able to transcribe polyphonic music
- Developed *Spectrum Subtraction Algorithm* to estimate Pitch and Amplitude.
  - Efficient: linear in number of pitches
- (Not shown) Extended to multi-instrument transcription.
Critique

How was the talk in terms of
Content
Audience
Medium  ?

How can it be improved?
**Summary**

Talk ≠ compress(paper)

- **Do Research** → **Tell a Story** → **Paper**
- **Talk** → **Re-telling Same Story**
- **Different versions**