11-823 Conlanging

Prosody 2: so what does it all mean?
Prosody

- **Timing**
  - Stress timed vs Syllable timed

- **Accents/stress**
  - Lexical/phonetic

- **Emotion/style**

- **Semantics and intonation**
Stress vs Syllable Timed

- **Length of Syllables**
  - equal(ish) or not
  - duh duh duh duh duh duh
  - duh duh DUH duh DUH duh DUH DUH

- **Syllable timed languages**
  - (Approximately) equal time for syllables
  - French, Japanese, Brazilian

- **Stress timed languages**
  - (sort of) equal time between stressed syllables
  - English, German, Portuguese
Other Timing Options

- **For conlanging:**
  - **Accent Groups/Intonational Phrases**
  - **equal(ish) timing**
  - **F0 range for stress**
  - **F0 range for sub-phrasing**
  - **Timing for sub-phrasing**
Lexical/Phonetic Stress/Tone

- **Change words with different prosody**
  - project(n) vs project(v)
  - 橋 (hashi' bridge) vs 筷 (ha'shi chopsticks)
  - 媽 (mā mother) vs 馬 (mǎ horse)

- **Wrong stress makes it hard to understand**
  - Oregano, address
  - THE girl IN THE park with the teLEScope
What is emotional speech

The standard 4 emotions
Neutral, Happy, Sad and Angry

But there are many more
Cold-anger, dominant, passive, shame
Confident, non-confident etc
English LDC Emotion (4 Emotions)

Short, 1-2 second, wav files
English speech – dates such as “November 3rd”
4 fundamental, distinct emotions
74 unique workers and 169 total HITs completed

Results

<table>
<thead>
<tr>
<th>Emotion</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>69%</td>
</tr>
<tr>
<td>Sadness</td>
<td>67%</td>
</tr>
<tr>
<td>Neutral</td>
<td>66%</td>
</tr>
<tr>
<td>Happiness</td>
<td>46%</td>
</tr>
<tr>
<td>Total</td>
<td>60%</td>
</tr>
</tbody>
</table>

Uni-directional Confusion
English LDC Emotion (15 Emotions)

Same parameters as previous experiment. Including less well-defined emotions Pride, shame, etc.

68 unique workers and 218 total HITs completed

### Results

<table>
<thead>
<tr>
<th>Emotion</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>29%</td>
</tr>
<tr>
<td>Hot-Anger</td>
<td>26%</td>
</tr>
<tr>
<td>Sadness</td>
<td>25%</td>
</tr>
<tr>
<td>Boredom</td>
<td>17%</td>
</tr>
<tr>
<td>Panic</td>
<td>14%</td>
</tr>
<tr>
<td>Interest</td>
<td>12%</td>
</tr>
<tr>
<td>Elation</td>
<td>10%</td>
</tr>
<tr>
<td>Contempt</td>
<td>10%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Emotion</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>9%</td>
</tr>
<tr>
<td>Pride</td>
<td>9%</td>
</tr>
<tr>
<td>Despair</td>
<td>8%</td>
</tr>
<tr>
<td>Cold-Anger</td>
<td>7%</td>
</tr>
<tr>
<td>Anxiety</td>
<td>5%</td>
</tr>
<tr>
<td>Disgust</td>
<td>5%</td>
</tr>
<tr>
<td>Shame</td>
<td>4%</td>
</tr>
<tr>
<td>Total</td>
<td>12%</td>
</tr>
</tbody>
</table>

Uni-directional Confusion
German Berlin Emotion (7 Emotions)

Short sentences with no emotional connotation

“The tablecloth is lying on the fridge.”

37 unique workers and 245 total HITs completed

Results

<table>
<thead>
<tr>
<th>Emotion</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>68%</td>
</tr>
<tr>
<td>Anger</td>
<td>62%</td>
</tr>
<tr>
<td>Sadness</td>
<td>53%</td>
</tr>
<tr>
<td>Anxiety</td>
<td>45%</td>
</tr>
<tr>
<td>Happiness</td>
<td>35%</td>
</tr>
<tr>
<td>Boredom</td>
<td>27%</td>
</tr>
<tr>
<td>Disgust</td>
<td>11%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>41%</strong></td>
</tr>
</tbody>
</table>

Common Confusion Pair

Boredom

Neutral

41.8%
Conversational Prosody

- We change our intonation on context
Conversational Variation

Base vs Conversational Recordings

Base: “Okay”

Conversational: “Okay”

Variation over same prompt

Different levels of apology etc
Select 21 dialogs

“Best” coverage of dialog acts
Record 795 utterances

Recording in a dialog

“User” is just a synthesizer
“User” is recorded actual user
“User” is a human

Dialog with human is more natural
3 speech databases

General speech (ARCTIC prompts)
1128 utts, 51 mins, isolated utts from novels

Let’s Go domain speech
2138 utts, 2hr16mins
Isolated Let’s Go domain utts

Let’s Go conversational
795 utts, 59mins
Recorded in dialog with human “user”
<table>
<thead>
<tr>
<th></th>
<th>F0D</th>
<th>MCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arctic</td>
<td>12.685</td>
<td>5.685</td>
</tr>
<tr>
<td>LetsGo</td>
<td>9.088</td>
<td>4.952</td>
</tr>
<tr>
<td>LetsGoC</td>
<td>12.531</td>
<td>5.192</td>
</tr>
</tbody>
</table>
Conversational Responses

- Answering “alright” in two different situation
- Transcript
  - User: 5am
  - System: Leaving at 5am. Did I get that right?
  - User: Yes
  - System: Alright
- Situation: Final state. After yes.
- Tone: Terminating, delight
Answering “alright” in three different situation

Transcript

- System: Okay. Penn Hills. Did I get that right?
- User: No
- System: Alright
- Situation: After no.
- Tone: Disappointment.
Two occurrence of “Is this correct”

Transcript

System: What can I do for you?
User: (silent noise).
System: The 1C. Is this correct?
User: (silent yes)
System: The 1C. Is this correct?

Situation: After no response. Repeated similar question.
Tone: Delightful → Demanding
Two occurrence of “Did I get that right?”

Transcript

S: 61A leaving East Pittsburgh tonight?
U: The 61A. Did I get that right?
S: Going to Swissvale?
U: The 61A. Did I get that right?

Situation: After unrelated respond. Repeated similar question.

Tone: Delightful → Demanding
**Style**

- **Styles:**
  - Formal, informal
  - Performed, Conversational

- **Genres**
  - Didactic, Politics, Humor
Simple Intonation Use

- **Contrast/Focus**
  - John saw Mark?
  - BILL saw Mark

- **Lists**
  - Strawberry, Apple, Banana.
  - Strawberry, Apple, Banana, ...

- **Question/Declarative (ish)**
  - John saw Mary?
  - John saw Mary.

- **Confidence**
  - Traveling to downtown, when will you leave
  - Traveling to downtown, when will you leave
More Complex Intonation

- Restrictive Relative clauses
  - The Swiss who like chocolate
  - Subset of all Swiss
  - The Swiss, who like chocolate
  - All Swiss (because they all like chocolate)
First/Second mention

- We intonationally reduce second mentions
  - The man saw the boy in the park.
  - The man gave him an ice cream
(non-)Verbal markers

- **Fillers**
  - *Uhm, em, eto, ano*
  - *Well, so*

- **Hesitations, false starts**
  - *Superfluous introductions …*

- “*like”, “you know”*

- **Cross lingual use**
Prosody

- Phrasing, Intonation, Duration, Power
- Intonational Phonology
  - Accent types
  - F0 generate
- Lexical Intonation
  - Tones, stress, lexical accent
  - Combinations (Tone Sondhi, Fudge Rules)
- Pragmatics
  - How people use intonational variants