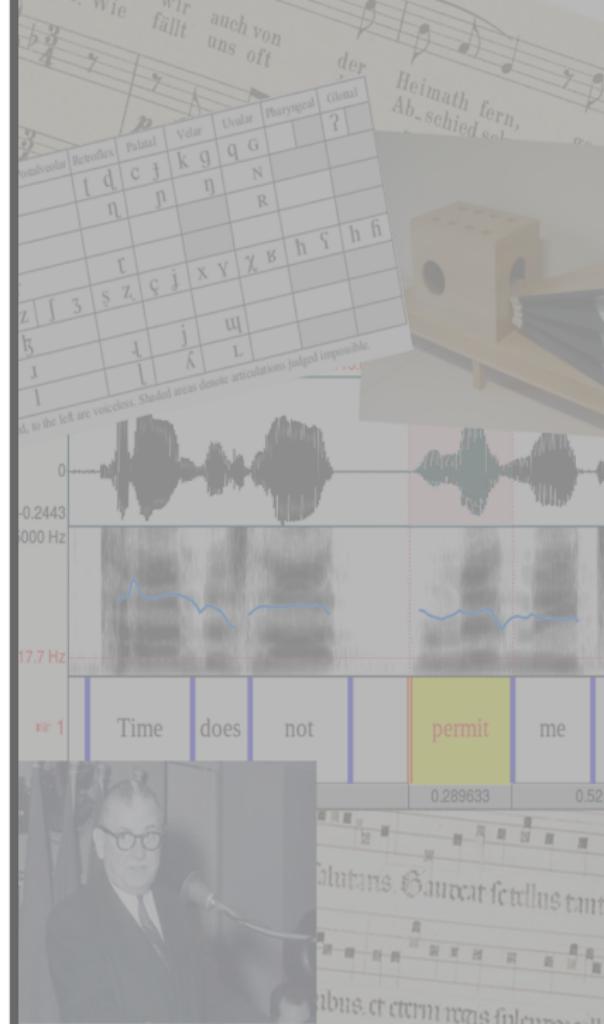


The MusLing Project

Investigating Joint Structures of Music and Linguistic Prosody

Emily K. Jamison

November 30, 2016



Overview

Introduction

Prosody: Speech Synthesis

 Introduction

 History

 Problems

Music (Western Tonal Music)

 Introduction

 History

 Fundamental questions

Combining music and speech prosody

 Initial evidence

 Labeling SD $\hat{1}$

 Predicting pauses

 Predicting scale degrees

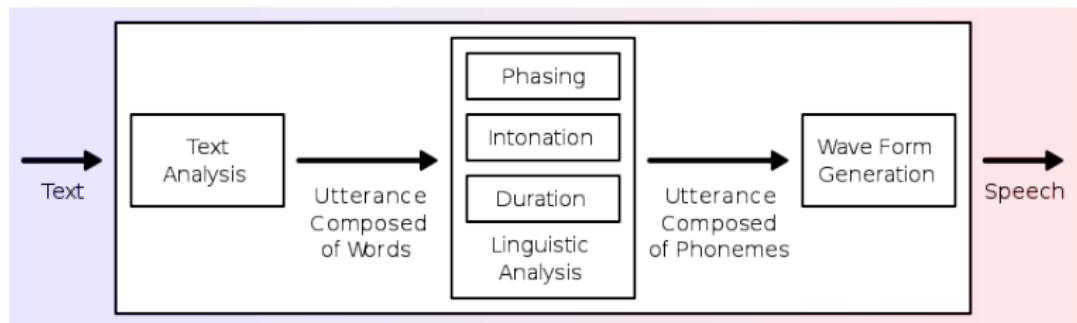
Future work

What is Speech Synthesis?

Speech Synthesis: the artificial production of human speech

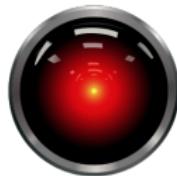
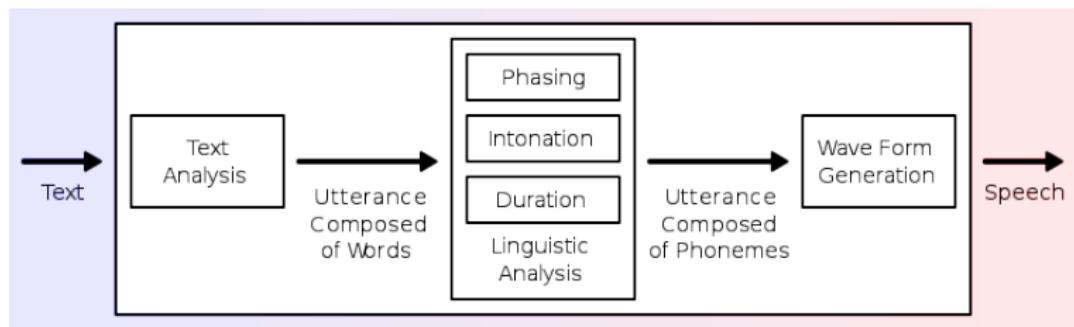
What is Speech Synthesis?

Speech Synthesis: the artificial production of human speech



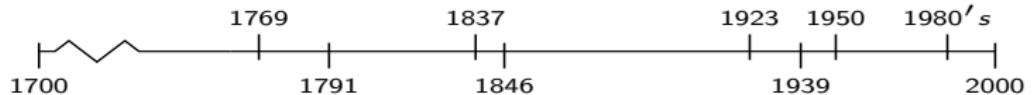
What is Speech Synthesis?

Speech Synthesis: the artificial production of human speech

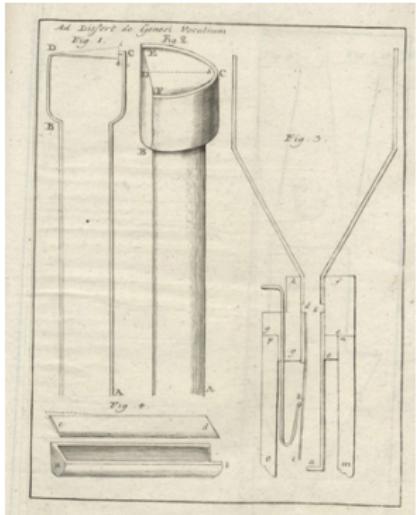


Google: "It's beautiful outside today, isn't it?"

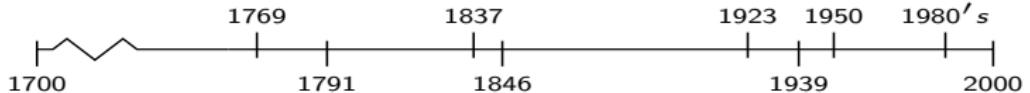
History of Speech Synthesis



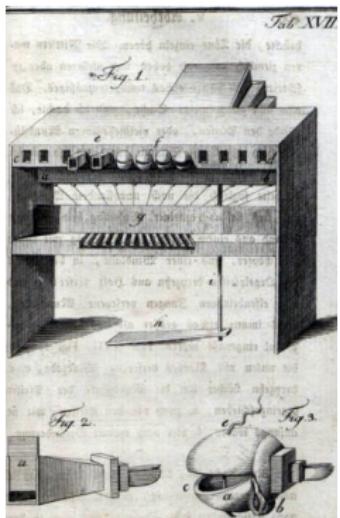
History of Speech Synthesis



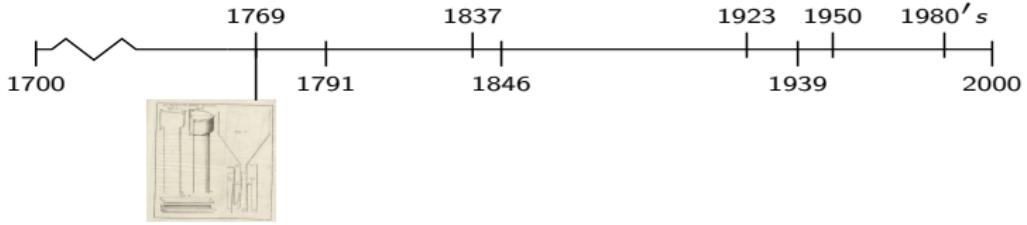
1779



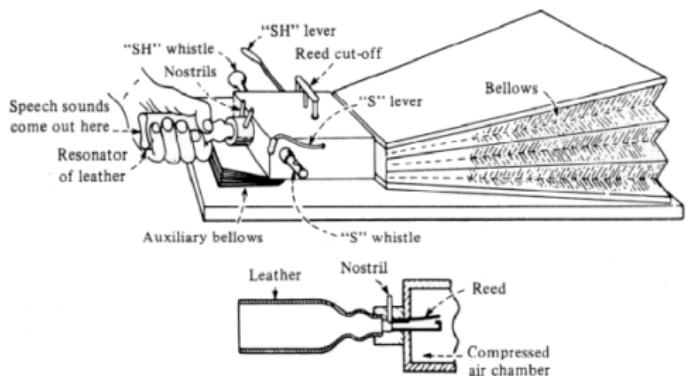
History of Speech Synthesis



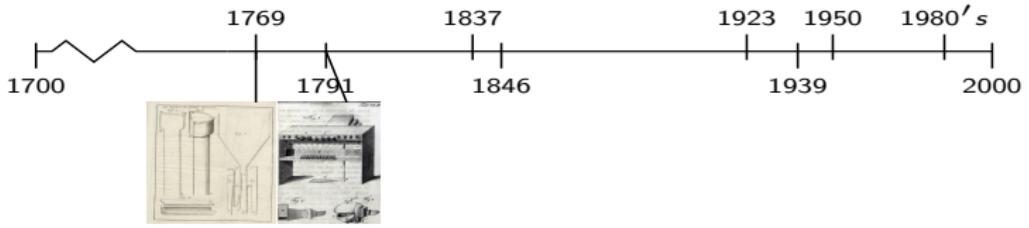
1791



History of Speech Synthesis



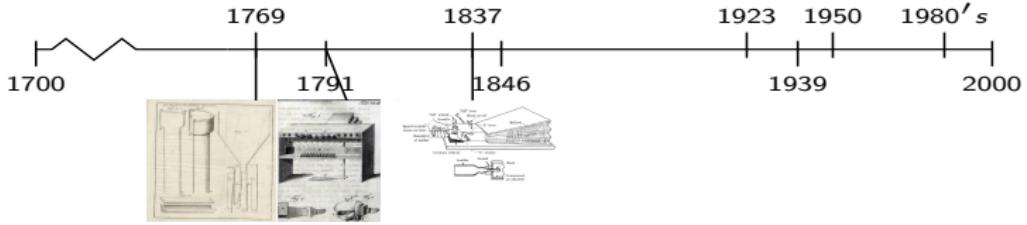
1837



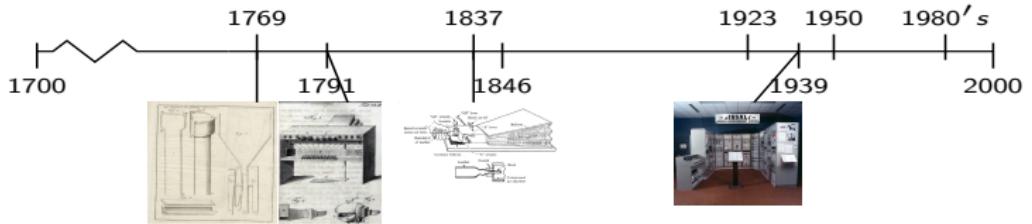
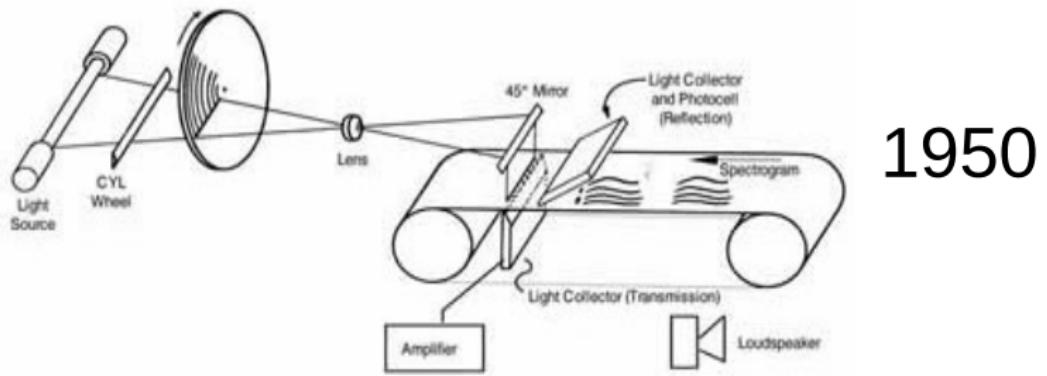
History of Speech Synthesis



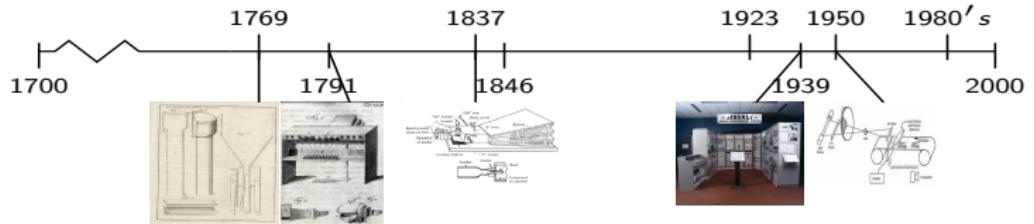
1939



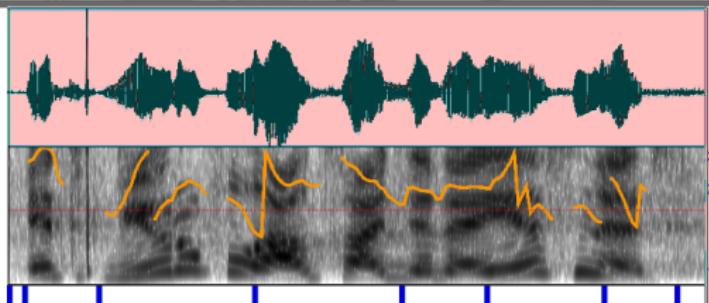
History of Speech Synthesis



History of Speech Synthesis



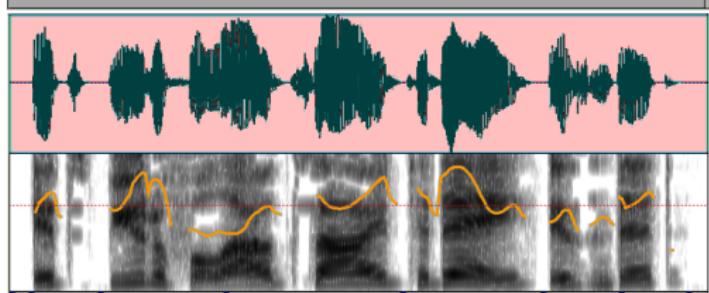
Spectrograms: Human versus Synthesizer



0167851

Visible part 1.711669 seconds

1.879520

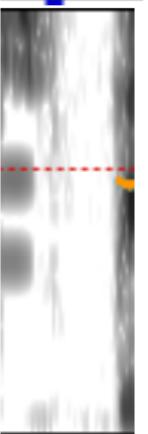
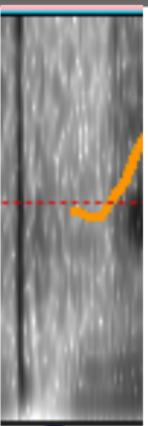
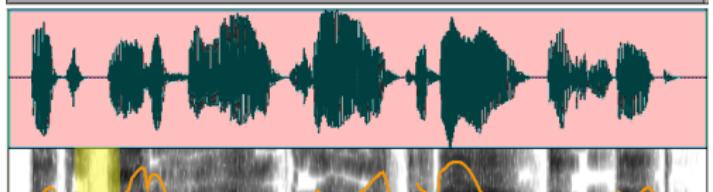
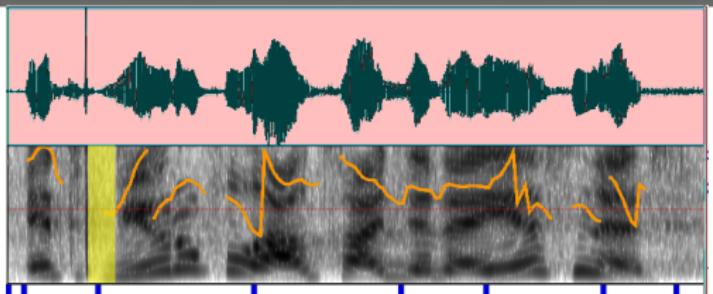


| It's

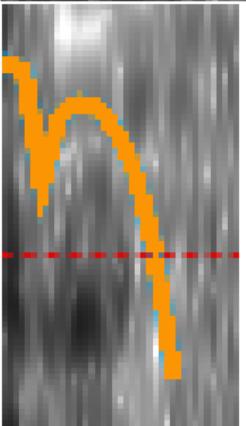
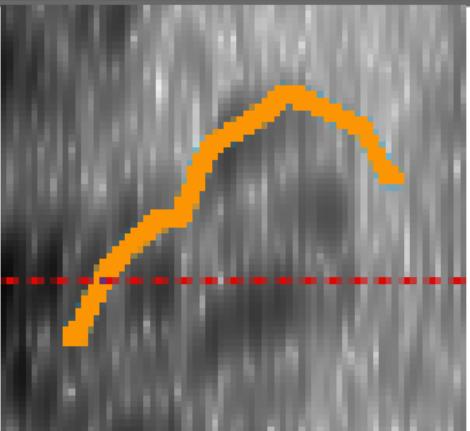
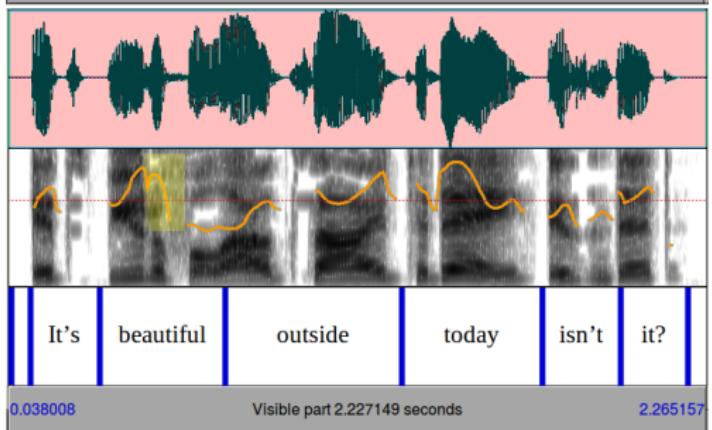
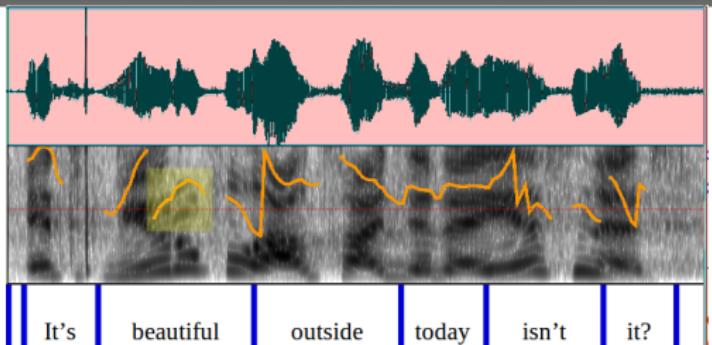
Visible part 2.227149 seconds

2.265157

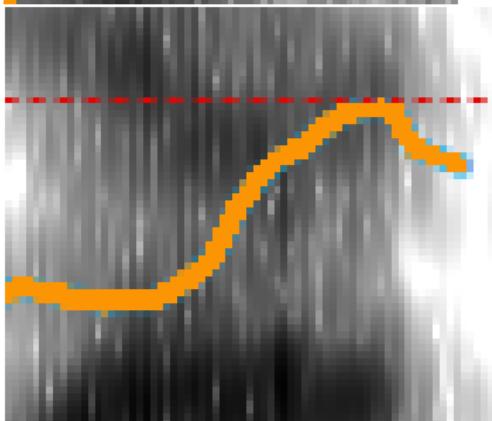
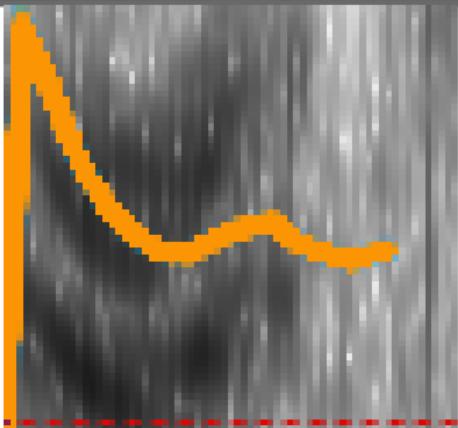
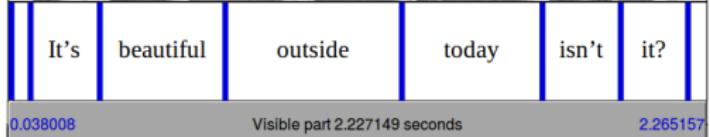
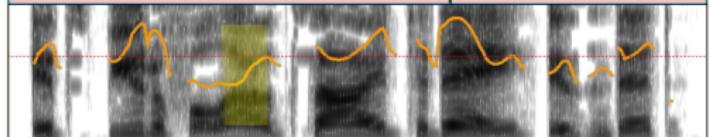
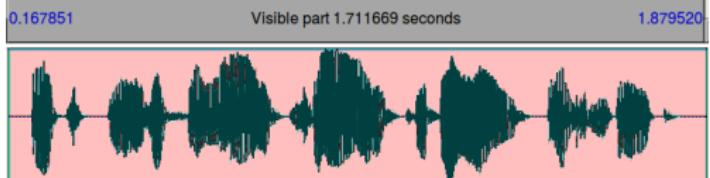
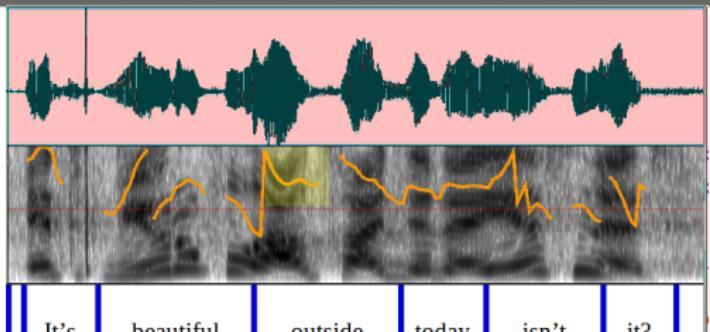
Spectrograms: Human versus Synthesizer



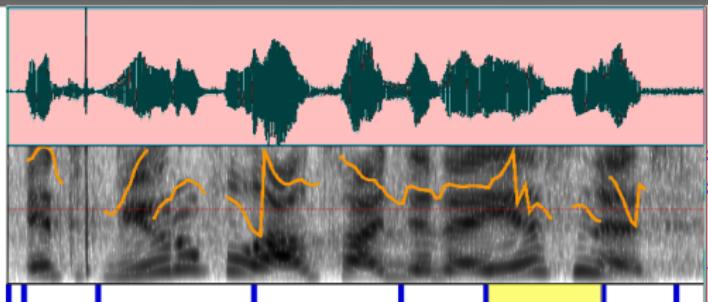
Spectrograms: Human versus Synthesizer



Spectrograms: Human versus Synthesizer



Spectrograms: Human versus Synthesizer



0.167851 Visible part 1.711669 seconds 1.879520



0.038008 Visible part 2.227149 seconds 2.265157

Overview

Introduction

Prosody: Speech Synthesis

 Introduction

 History

 Problems

Music (Western Tonal Music)

 Introduction

 History

 Fundamental questions

Combining music and speech prosody

 Initial evidence

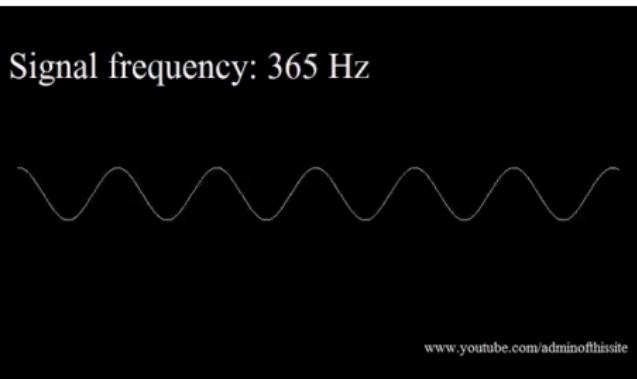
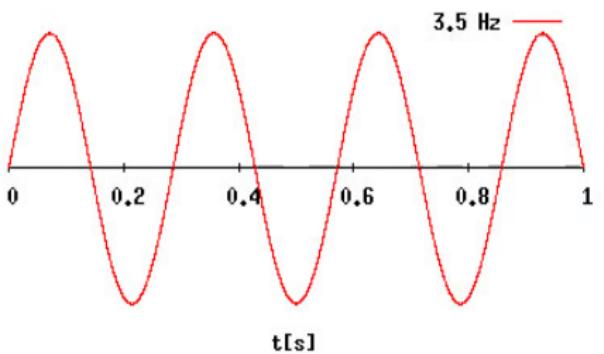
 Labeling SD $\hat{1}$

 Predicting pauses

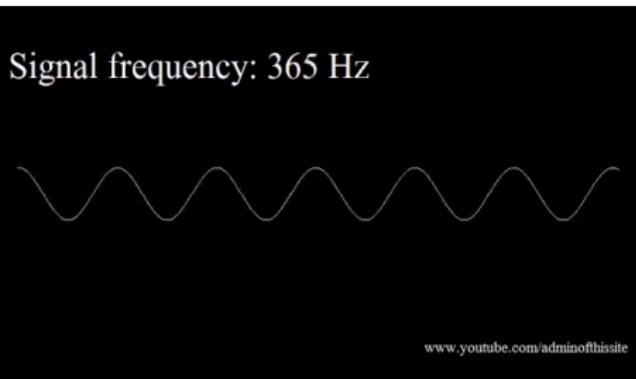
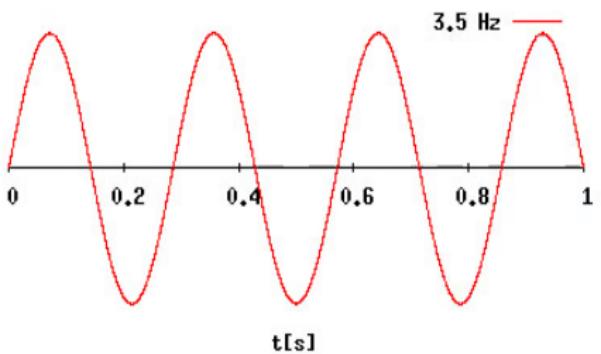
 Predicting scale degrees

Future work

Western Tonal Music

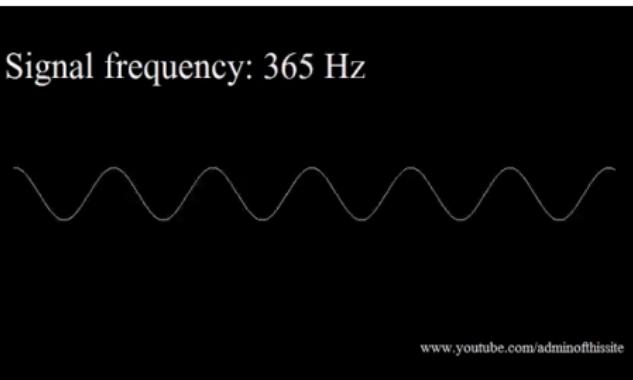
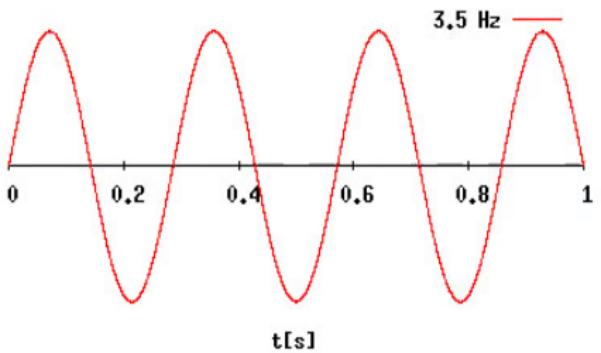


Western Tonal Music



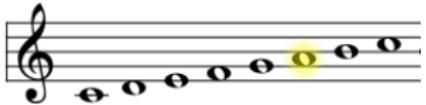
Example: 441 Hz is in interval [427.47 - 452.89]

Western Tonal Music

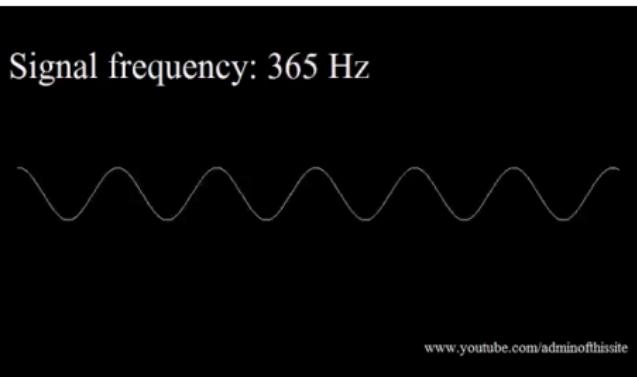
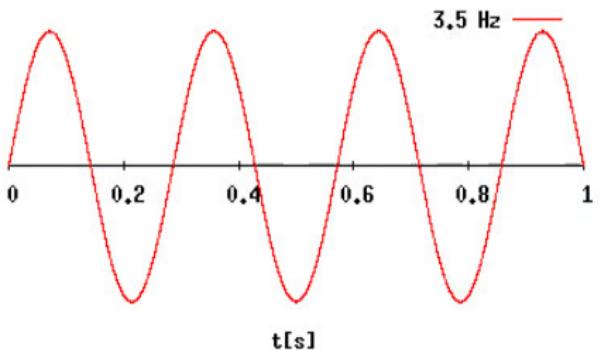


Example: 441 Hz is in interval [427.47 - 452.89]

[427.47 - 452.89] = musical note a

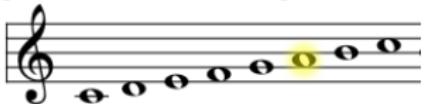


Western Tonal Music

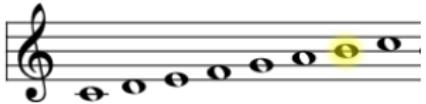


Example: 441 Hz is in interval [427.47 - 452.89]

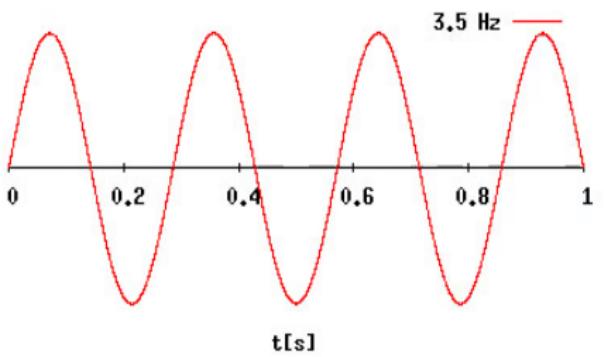
[427.47 - 452.89] = musical note *a*



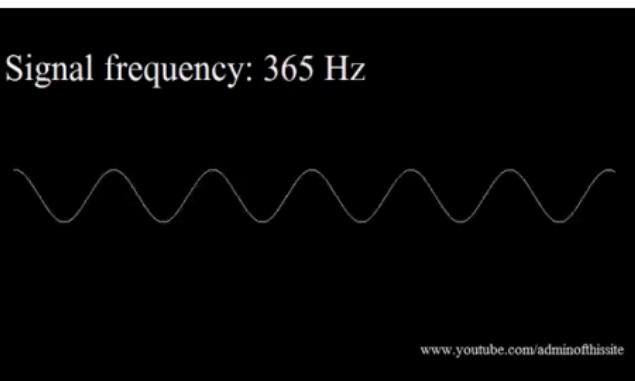
[479.82 - 508.35] = musical note *b*



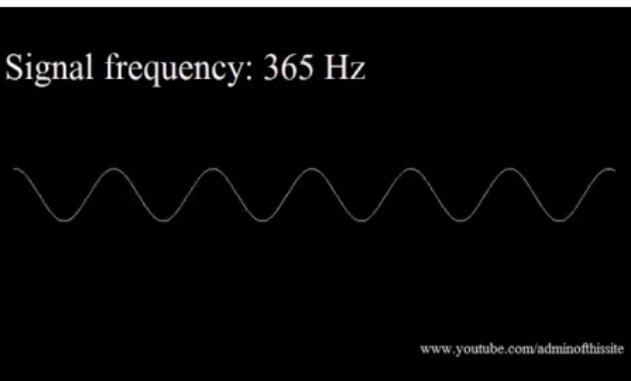
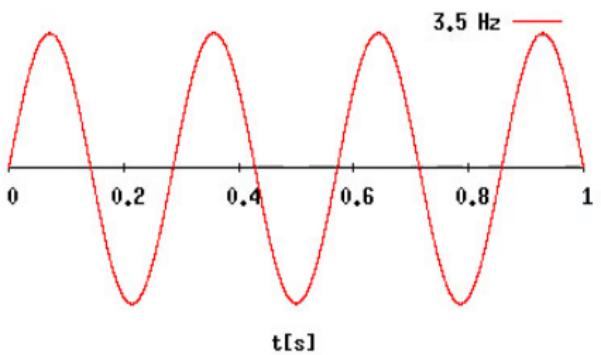
Western Tonal Music



Scale degrees: In C major, c is $\hat{1}$ and g is $\hat{5}$.



Western Tonal Music

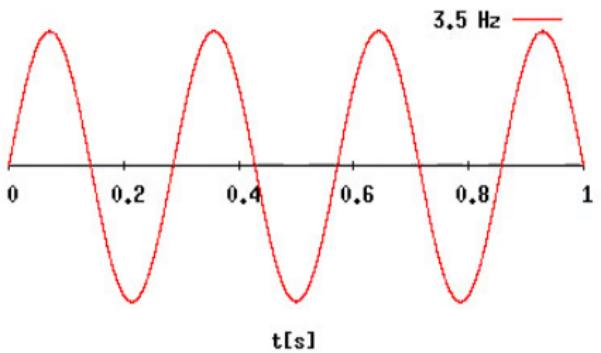


Scale degrees: In C major, c is $\hat{1}$ and g is $\hat{5}$.



In D major, d is $\hat{1}$ and a is $\hat{5}$.

Western Tonal Music

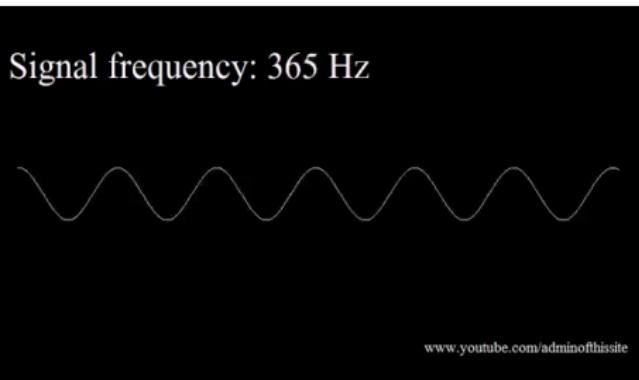


Scale degrees: In C major, c is $\hat{1}$ and g is $\hat{5}$.



In D major, d is $\hat{1}$ and a is $\hat{5}$.

Scale degrees: key equivalence. Key not required.



Rhythm and Accent in Music



$\text{♩} = 80-160$

A musical staff with a tempo marking of 80-160 BPM. The staff begins with a 4:4 time signature, indicated by a '4' above the staff. It then changes to a bass clef, followed by a 2:4 time signature, indicated by a '2' above the staff. The staff concludes with a repeat sign and a double bar line.

History of Western Tonal Music

Musical scales: ratio-defined sets of notes. E.g.
Eb melodic minor scale.

History of Western Tonal Music

Musical scales: ratio-defined sets of notes. E.g.
Eb melodic minor scale.

Musical modes: different interval ratios within a
musical scale E.g. Lydian, Phrygian

History of Western Tonal Music

Musical scales: ratio-defined sets of notes. E.g.
*E*b melodic minor scale.

Musical modes: different interval ratios within a musical scale E.g. Lydian, Phrygian

Harmony: the ratio between a simultaneous high note and low note

History of Western Tonal Music

Musical scales: ratio-defined sets of notes. E.g.
Eb melodic minor scale.

Musical modes: different interval ratios within a musical scale E.g. Lydian, Phrygian

Harmony: the ratio between a simultaneous high note and low note

Melos: categories of rhythmic composition

History of Western Tonal Music

Polyphony (< mid-1400's): synchronous lines of music

History of Western Tonal Music

Polyphony (< mid-1400's): synchronous lines of music

Species counterpoint: < 1532 AD

- pedagogic tool of music modeling
- strict set of rules and preferences
- controls both sequential and synchronous choice of note

One Note Against One Note
First Species

Only Harmonic Intervals Allowed Are:
1, 3, 5, 6, or 8 (*consonances*)

Interval must be 1, 5, or 8
(*Pitches must be Topic or Dominant*)

must be 1, or 8

interval: 5 6 6 3 3 6 6 8 6 6 6 8

type of motion: S P C P C P C C P P C Cadence

Contrary Motion most desirable.

Parallel Motion:
3, 3 OK (good);
6, 6 OK (good);
5, 5 BAD;
8, 8 BAD

Must Use Contrary Motion

What caused Western Tonal Music?

Where did the concept of music come from?

What caused Western Tonal Music?

Where did the concept of music come from?

What caused musical rules?

What caused Western Tonal Music?

Where did the concept of music come from?

What caused musical rules?

What is music?

Overview

Introduction

Prosody: Speech Synthesis

 Introduction

 History

 Problems

Music (Western Tonal Music)

 Introduction

 History

 Fundamental questions

Combining music and speech prosody

 Initial evidence

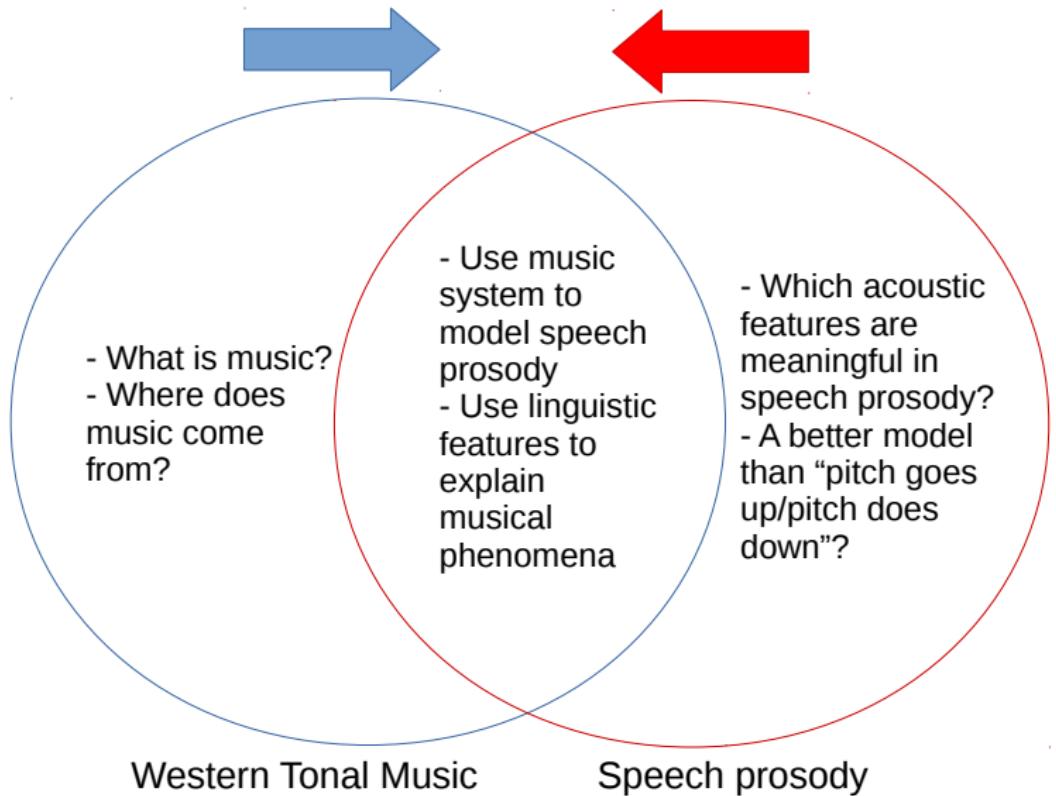
 Labeling SD $\hat{1}$

 Predicting pauses

 Predicting scale degrees

Future work

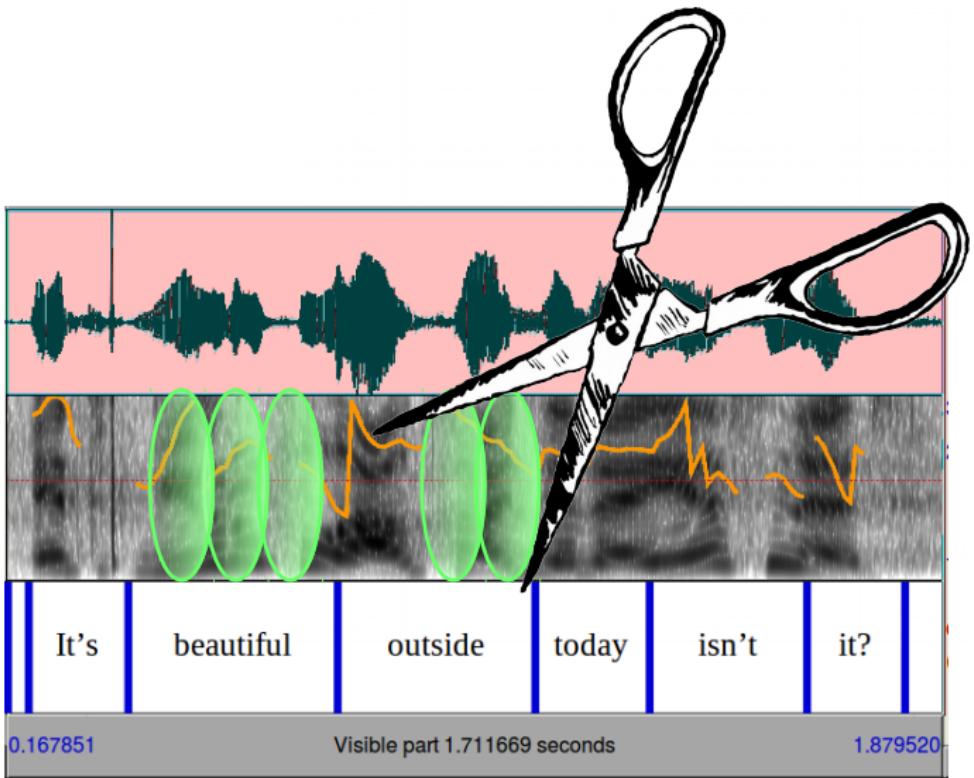
Our Research Topics



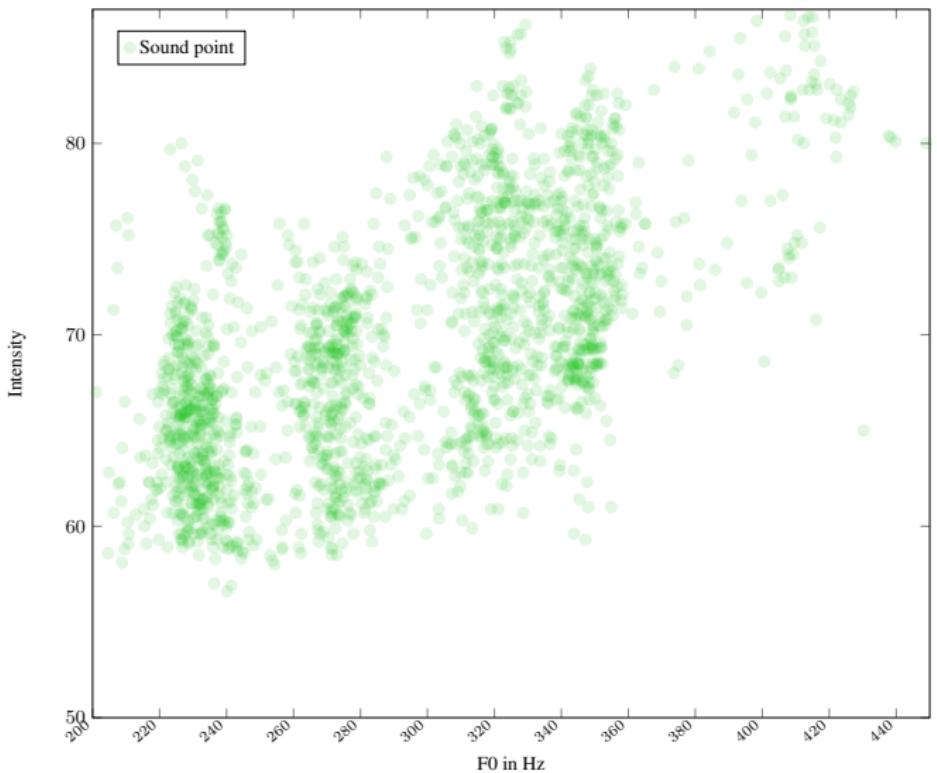
Some Research Questions

- Evidence for music-speech connection?
- Labeling $\hat{1}$?
- Predict pauses?
- Predict scale degrees?

Sampling F0

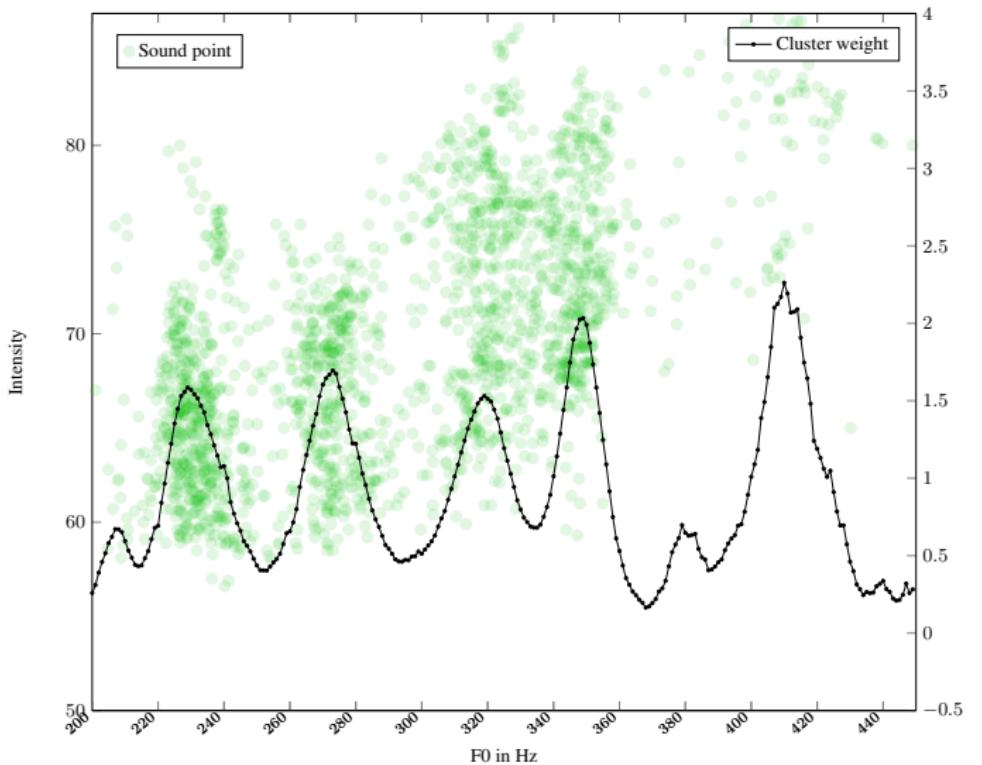


Rediscovering musical intervals



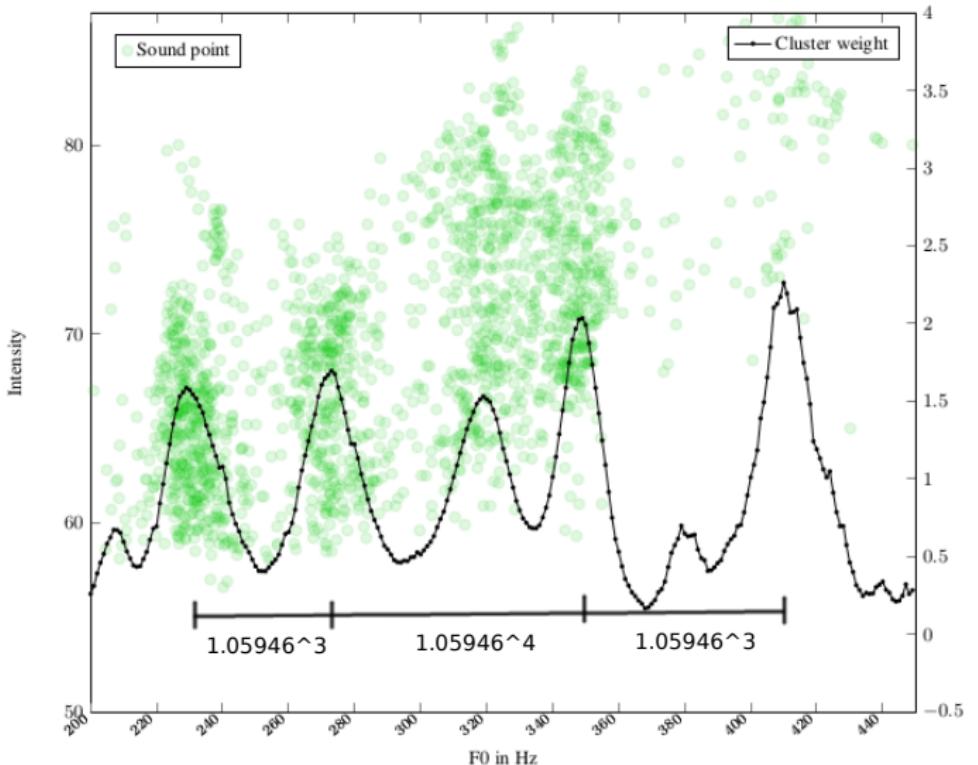
Adele, “Rolling in the Deep” from news interview.

Rediscovering musical intervals



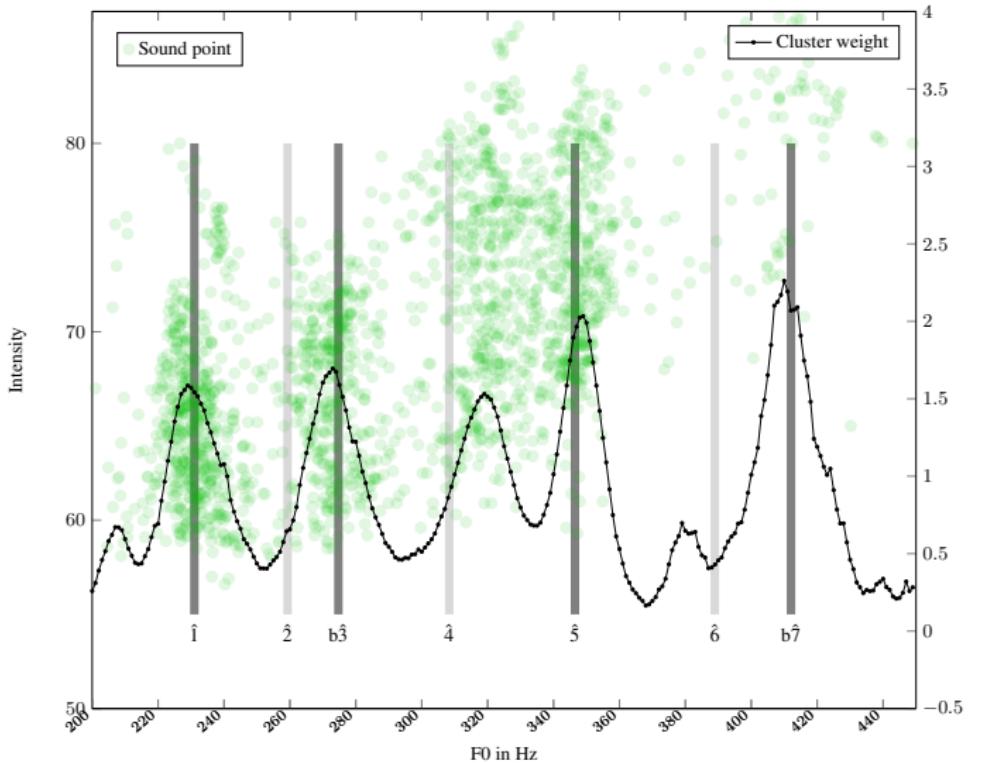
Adele, "Rolling in the Deep" from news interview.

Rediscovering musical intervals



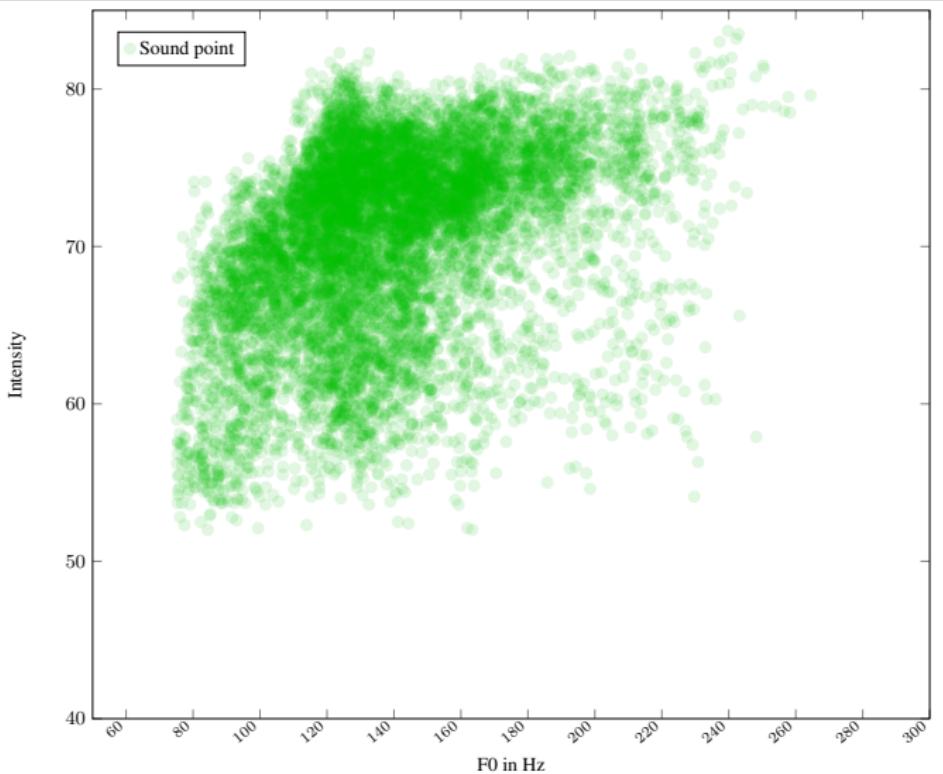
Adele, "Rolling in the Deep" from news interview.

Rediscovering musical intervals



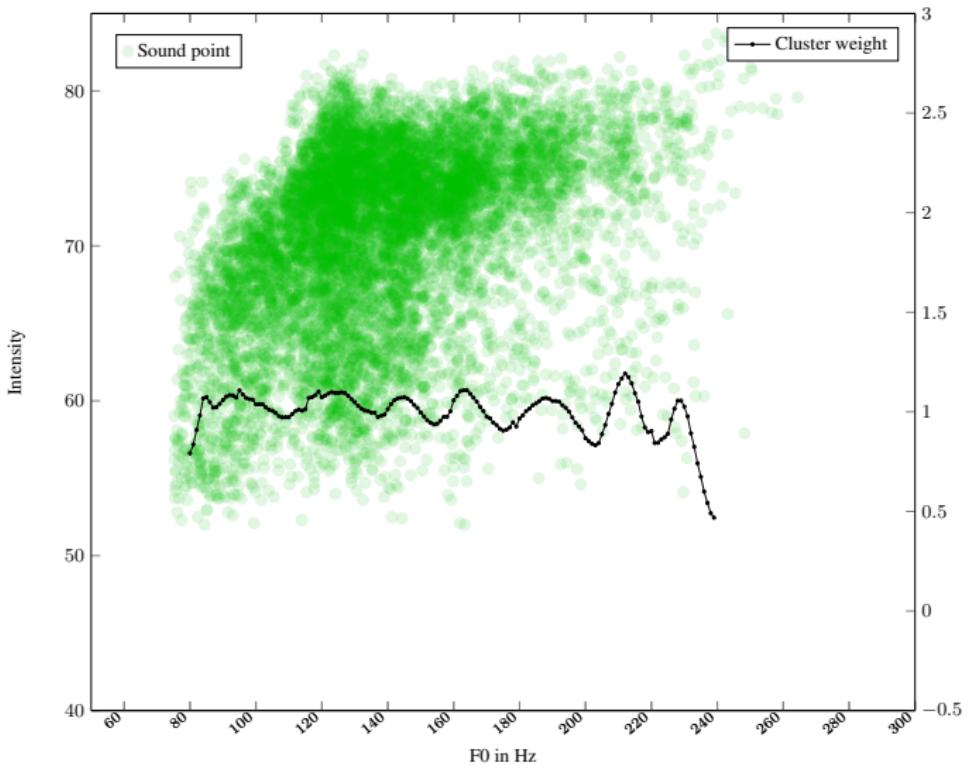
Adele, "Rolling in the Deep" from news interview.

Musical intervals in speech



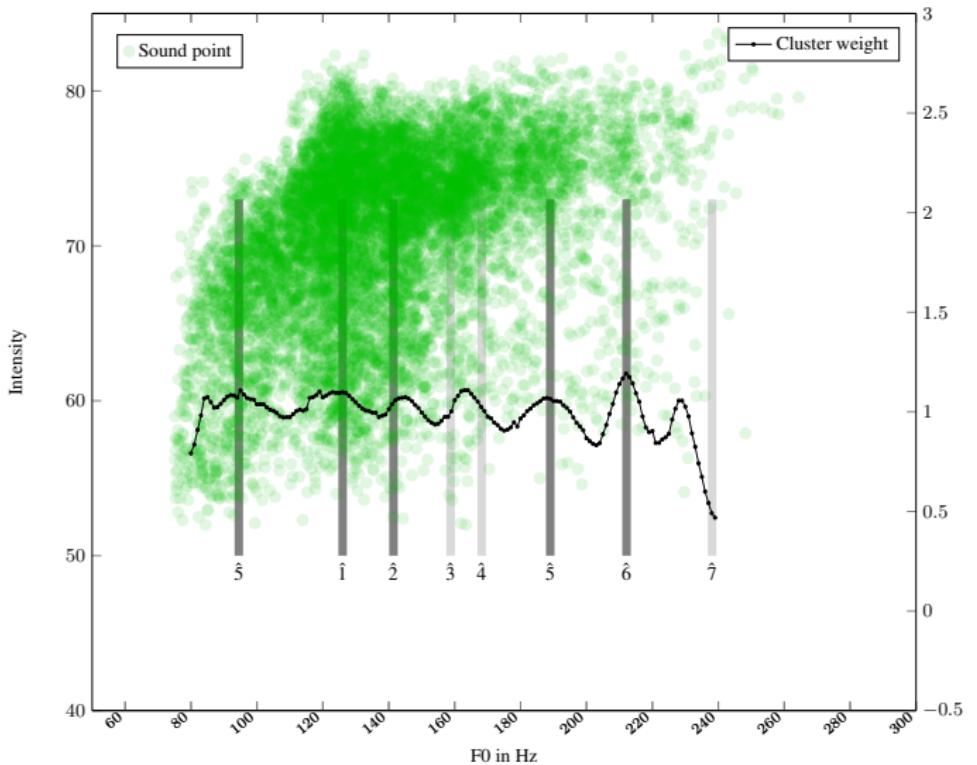
US President Dwight Eisenhower, Farewell Address, 1961

Musical intervals in speech



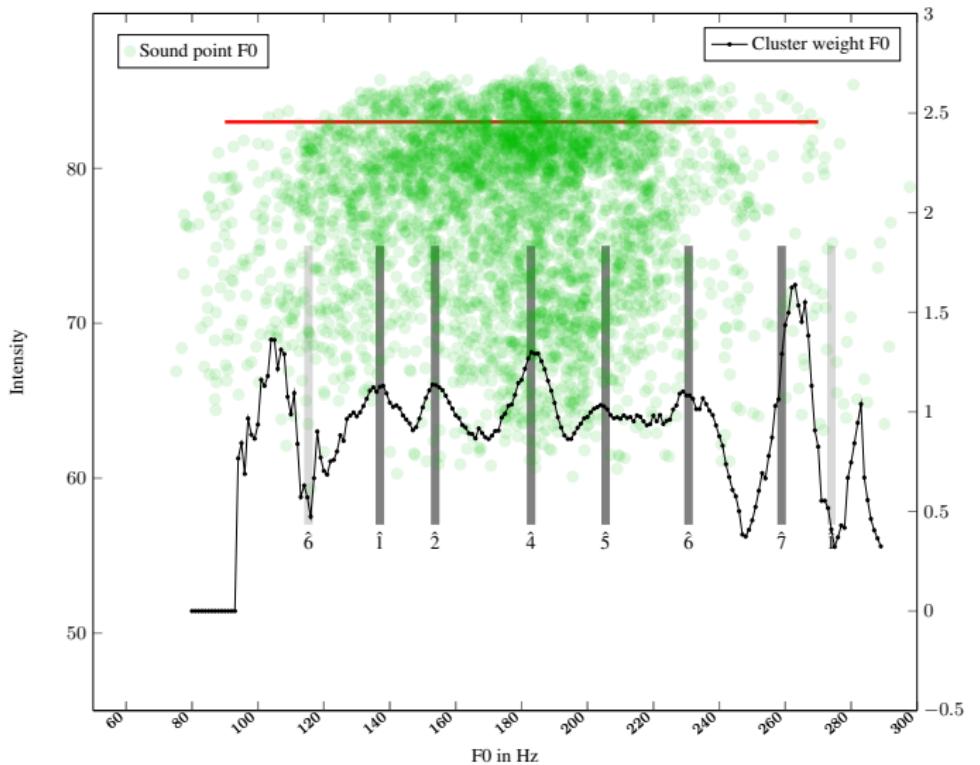
US President Dwight Eisenhower, Farewell Address, 1961

Musical intervals in speech



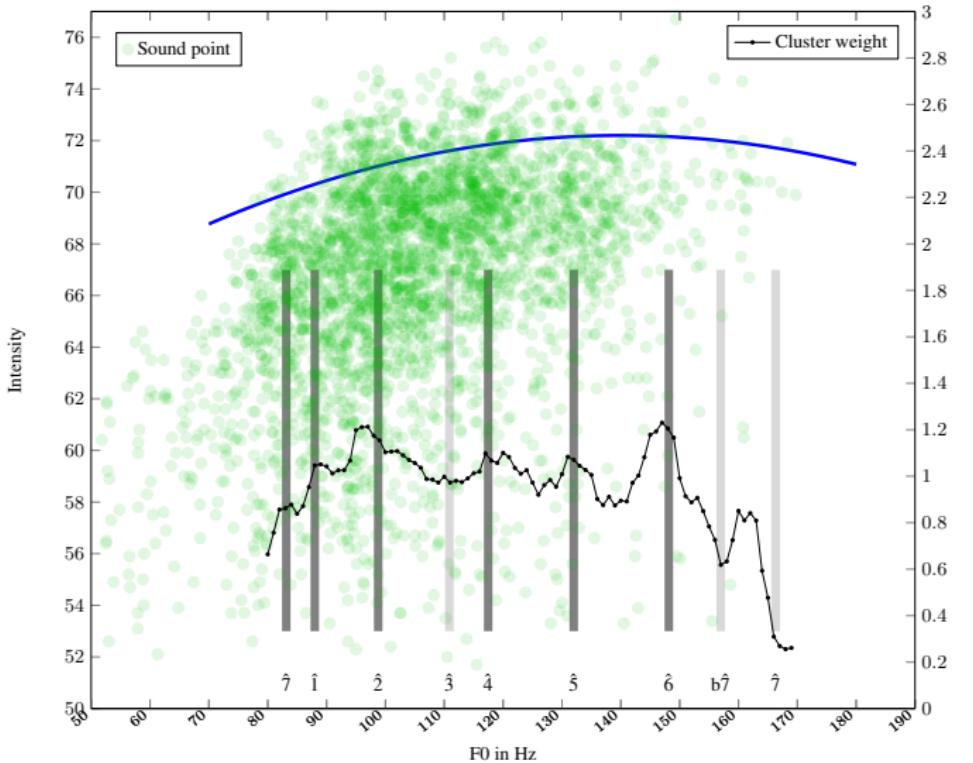
US President Dwight Eisenhower, Farewell Address, 1961

Musical intervals in speech



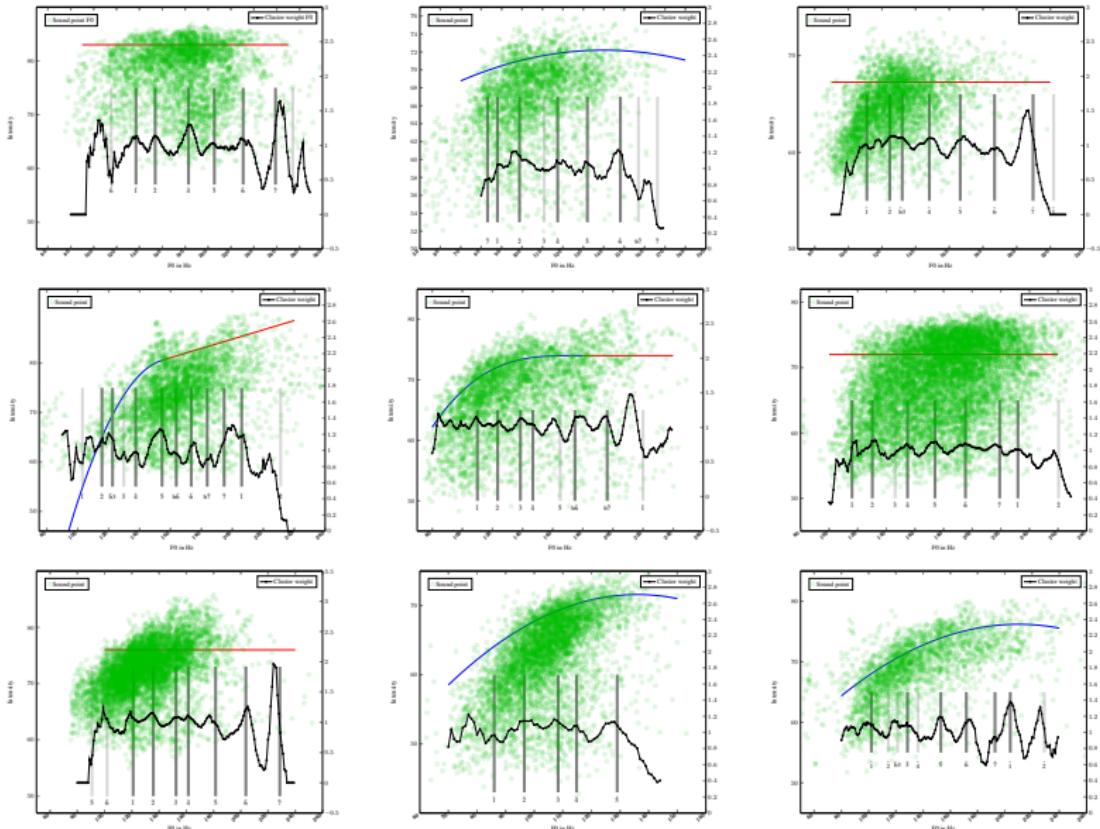
US president Franklin D. Roosevelt, Declaration of War, 1941

Musical intervals in speech

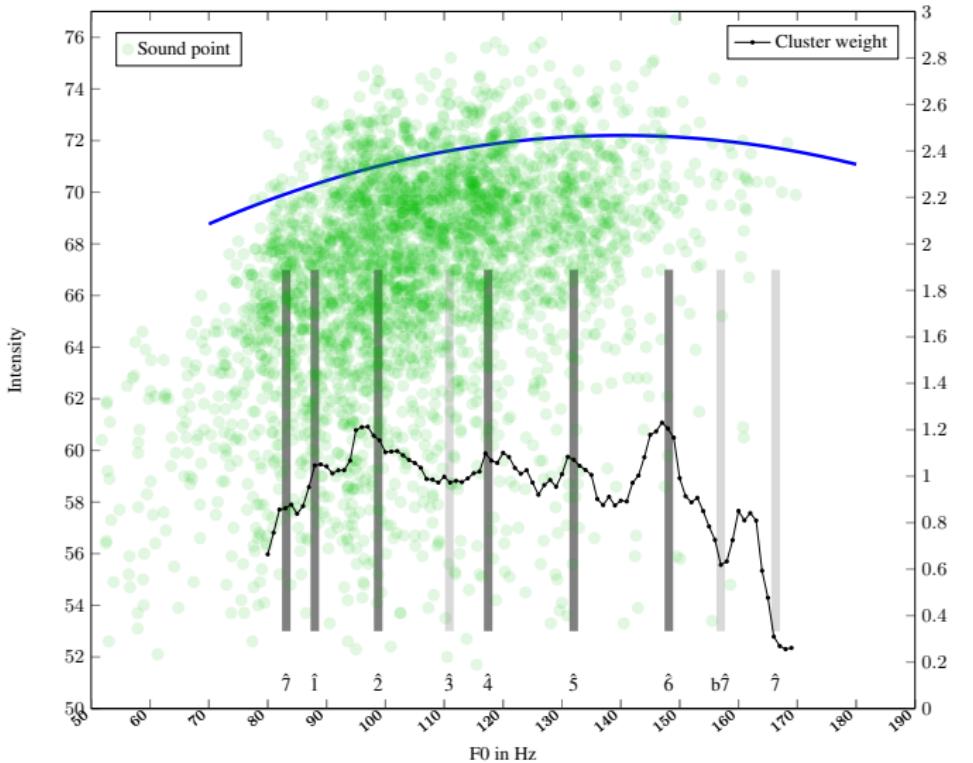


US president George Bush Sr., War on Iraq, 1991

Musical intervals in speech

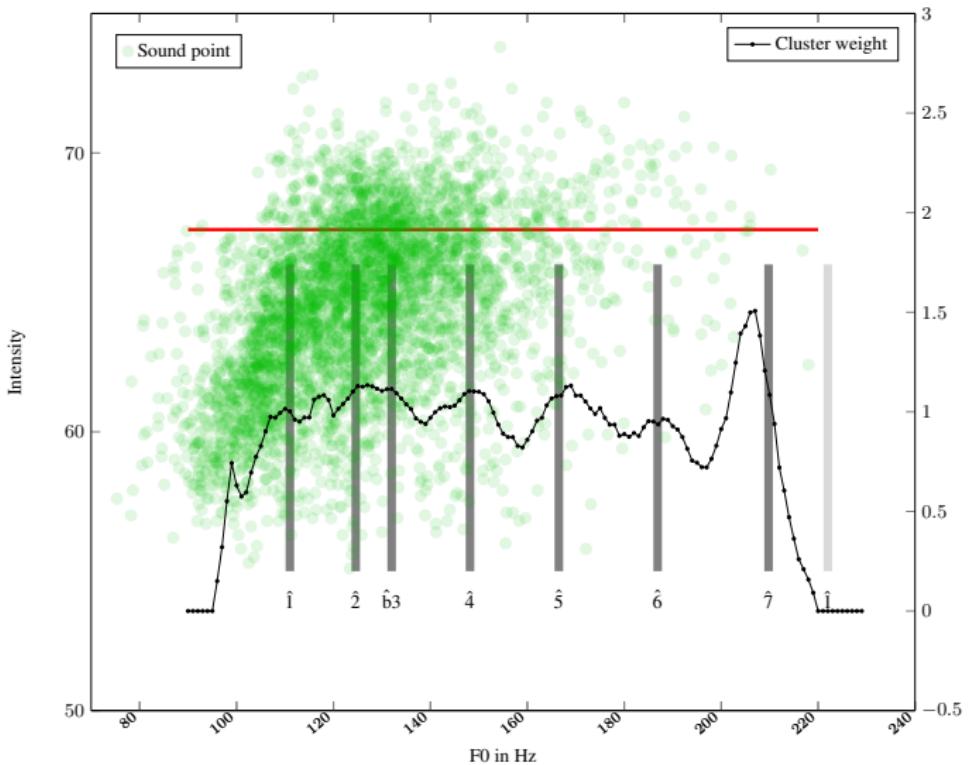


Musical intervals in speech



Goebbels speech, Third Reich, 1943

Musical intervals in speech



Hitler speech, Third Reich, pre-1941

Musical intervals in speech

Sequence = $\hat{1}, \hat{6}, \hat{5}, \hat{b}3, \hat{1}, \hat{5}$

Musical intervals in speech

Sequence = $\hat{1}, \hat{6}, \hat{5}, \hat{b}3, \hat{1}, \hat{5}$

Average of $P(\hat{6} | \hat{1}), P(\hat{5} | \hat{6}), P(\hat{b}3 | \hat{5}), P(\hat{1} | b\hat{3}), P(\hat{5} | \hat{1})$

Musical intervals in speech

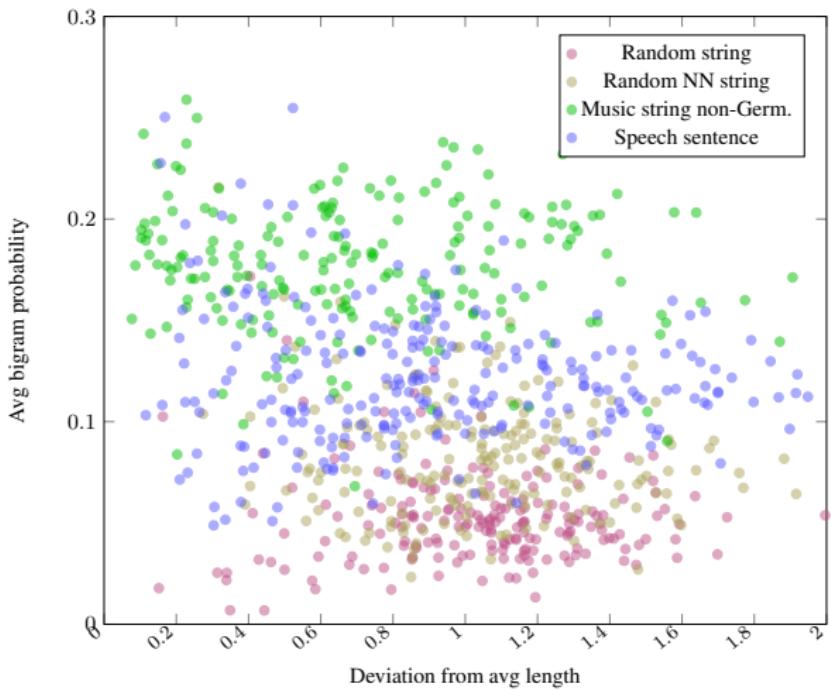
Random: b $\hat{7}$, $\hat{4}$, $\hat{3}$, b $\hat{5}$, # $\hat{4}$, b $\hat{4}$

Neighbor Notes: $\hat{4}$, # $\hat{3}$, # $\hat{2}$, b $\hat{1}$, $\hat{1}$, b $\hat{1}$

Music: b $\hat{7}$, $\hat{1}$, $\hat{5}$, b $\hat{6}$, $\hat{5}$, $\hat{4}$

Speech: b $\hat{3}$, $\hat{4}$, $\hat{3}$, b $\hat{3}$, b $\hat{6}$, $\hat{5}$

Musical intervals in speech



Speech note sequences are **not** random.

Musical intervals in speech

Observation: Aggregated speech and music have similarities.

Musical intervals in speech

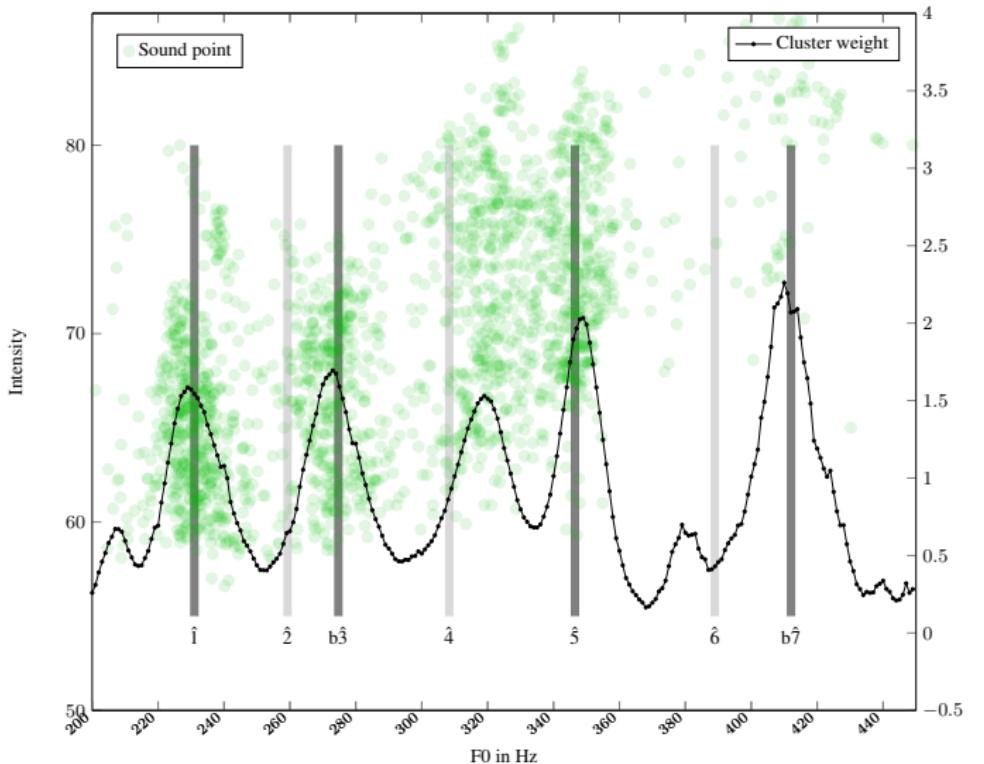
Observation: Aggregated speech and music have similarities.

Next: Investigate musical intervals at the linguistic level

Some Research Questions

- Evidence for music-speech connection?
- Labeling $\hat{1}$?
- Predict pauses?
- Predict scale degrees?

Finding SD1 in music



Adele, "Rolling in the Deep" from news interview.

Finding SD $\hat{1}$ in music

- Where is the tritone?
- Key signature plus accidentals
- Diatonic chord progression (i.e., subdominant → dominant → tonic)

Finding SD $\hat{1}$ in music

- Where is the tritone?
- Key signature plus accidentals
- Diatonic chord progression (i.e., subdominant → dominant → tonic)

Finding SD $\hat{1}$ in music

- Where is the tritone?
- Key signature plus accidentals
- Diatonic chord progression (i.e., subdominant → dominant → tonic)

Finding SD $\hat{1}$ in music

Experiment description	Ger.	Ital.	Fr.	Nor.	Amer.	all
# pieces	265	146	53	14	21	499
# composers	9	4	2	1	4	20
• All notes in all chords	.96	.91	.87	.86	.90	.93
• Only notes in new chords	.95	.91	.83	.71	.86	.91
• Soprano notes in all chords	.79	.77	.64	.79	.52	.76
• Soprano notes in new chords	.77	.69	.57	.64	.62	.72
• Only new soprano notes	.70	.64	.58	.71	.76	.67

Table: Percentage of pieces in which SD $\hat{5}$ or SD $\hat{1}$ is most frequent.

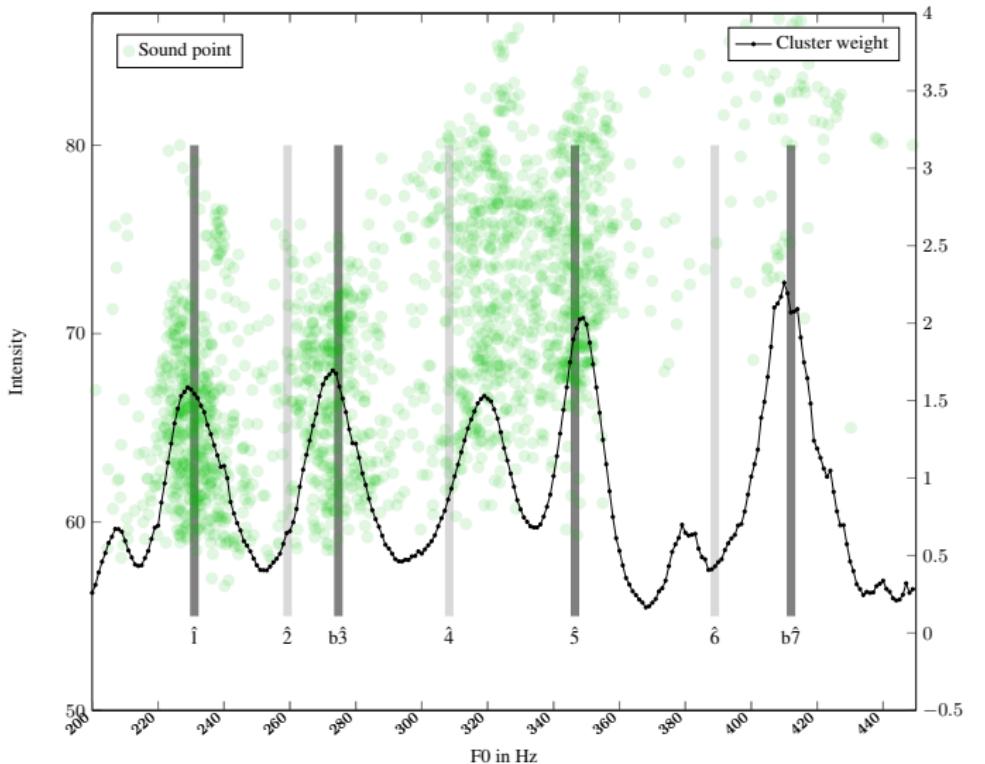
Finding $\hat{SD1}$ in music

Experiment description	Ger.	Ital.	Fr.	Nor.	Amer.	all
# pieces	265	146	53	14	21	499
# composers	9	4	2	1	4	20
• All notes in all chords	.96	.91	.87	.86	.90	.93
• Only notes in new chords	.95	.91	.83	.71	.86	.91
• Soprano notes in all chords	.79	.77	.64	.79	.52	.76
• Soprano notes in new chords	.77	.69	.57	.64	.62	.72
• Only new soprano notes	.70	.64	.58	.71	.76	.67

Table: Percentage of pieces in which $SD\hat{5}$ or $SD\hat{1}$ is most frequent.

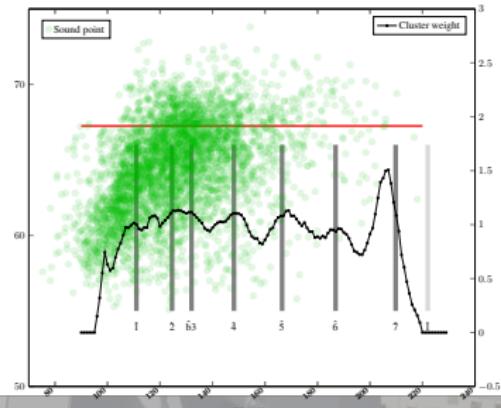
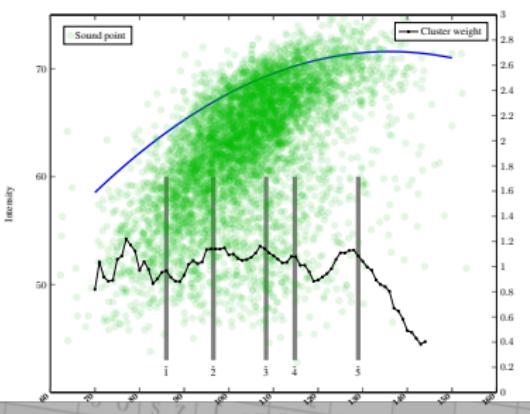
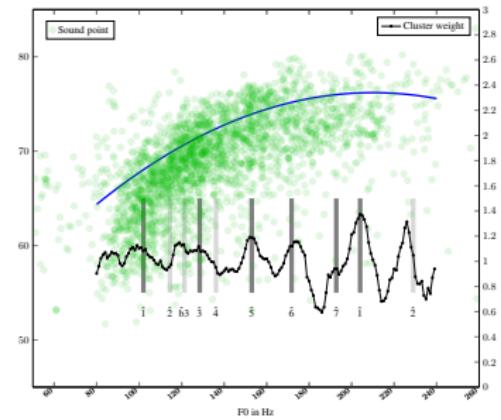
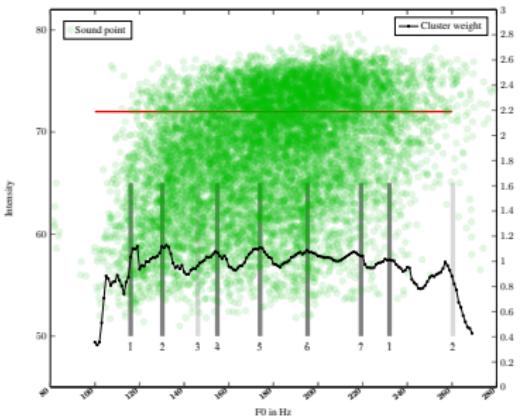
The densest scatterplot peak is probably $\hat{1}$ or $\hat{5}$.

Finding SD1 in music



Adele, "Rolling in the Deep" from news interview.

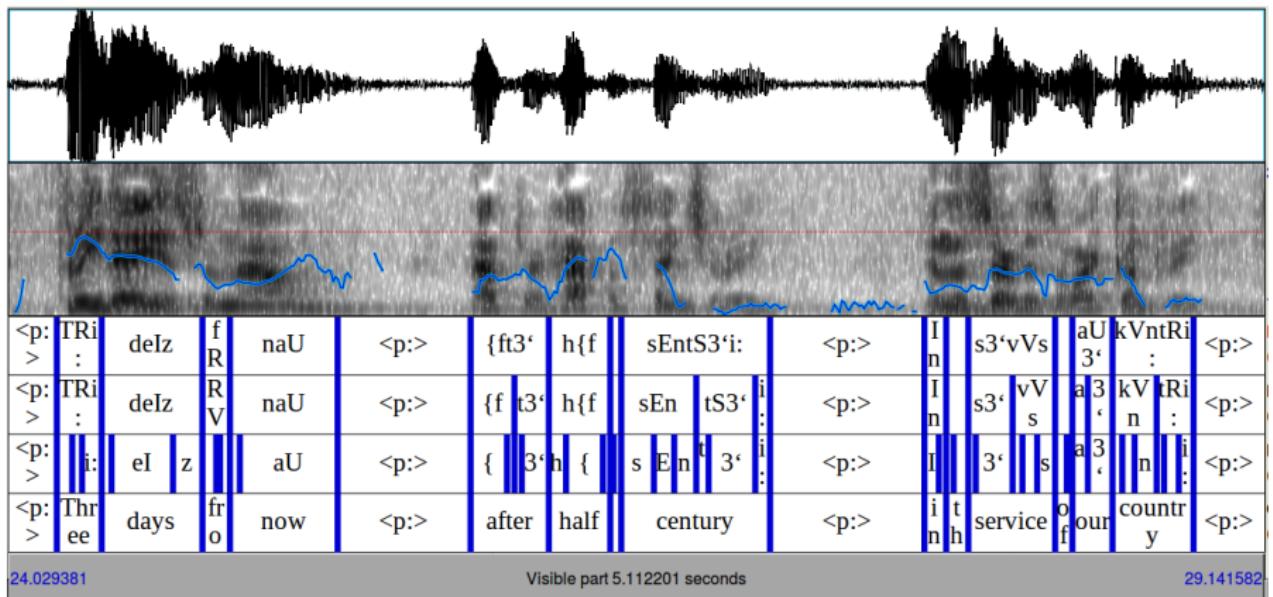
Finding SD1 in speech



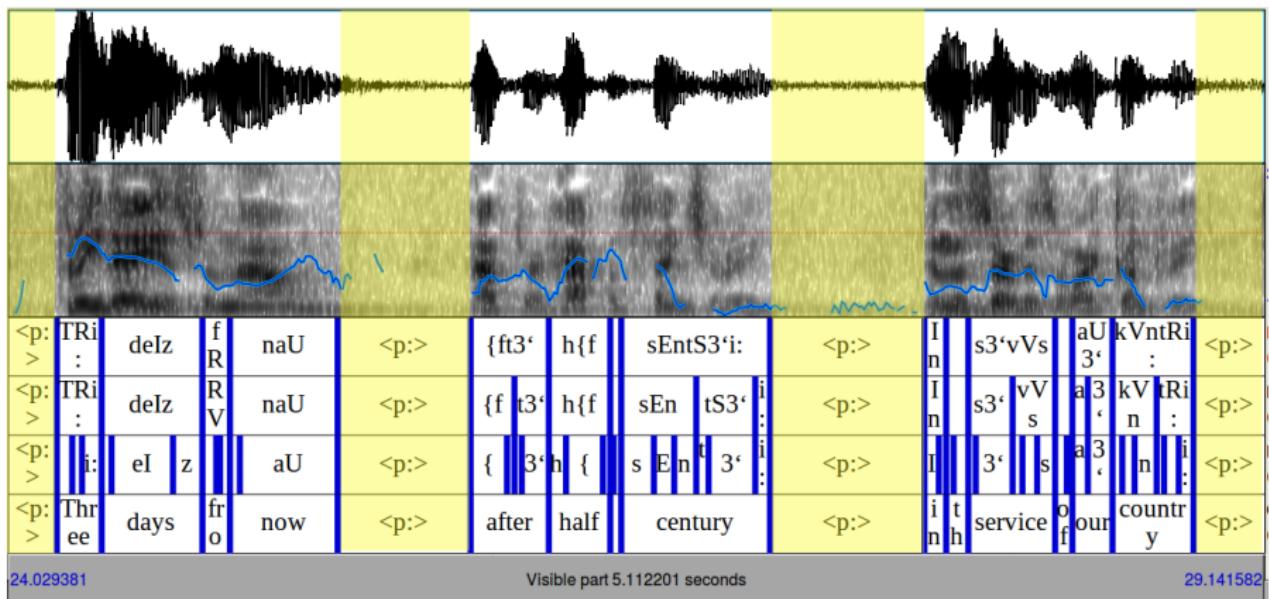
Some Research Questions

- Evidence for music-speech connection?
- Labeling $\hat{1}$?
- Predict pauses?
- Predict scale degrees?

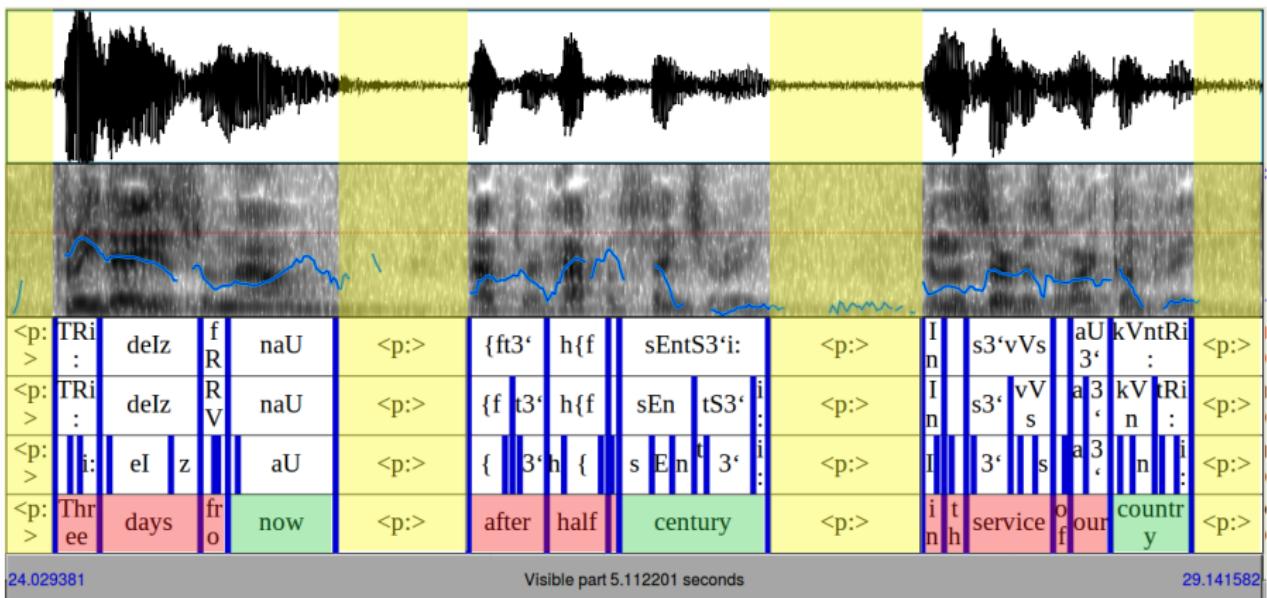
Predicting Pauses



Predicting Pauses



Predicting Pauses



Predicting Pauses

3 Feature types:

*Sentence: The **cat** chased the mouse.*

- word *Ex: 'cat'*
- POS-tag *Ex: 'NN'*
- set of syntactic tree kernels, each including parent and children nodes, with ancestor child marked. All ancestor kernels are used.
Ex: NP_Det-!NN, S_-!NP-VP

Predicting Pauses

Dataset 1 size: NoP=3485, P=771 (18%)

Features used	$F_1(\text{NoP})$	$F_1(\text{P})$
WORDS	.920	.426
WORDS POS	.919	.514
WORDS TK	.925	.597
WORDS POS TK	.927	.635

Dataset 2 size: NoP=3741, P=850 (19%)

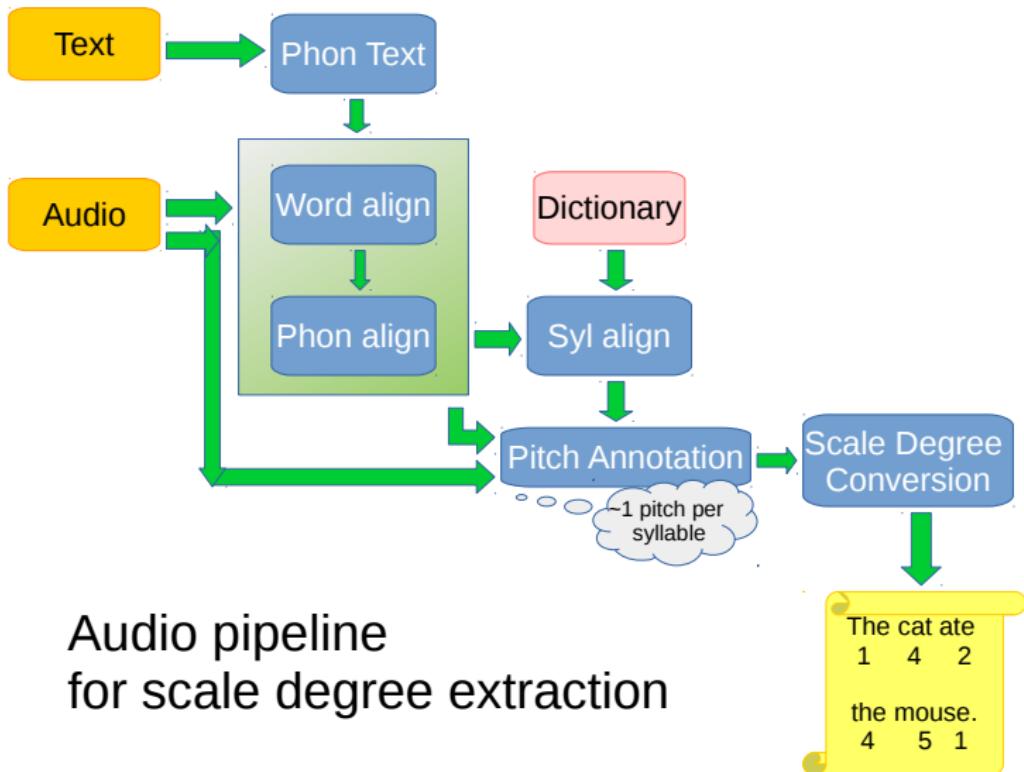
Features used	$F_1(\text{NoP})$	$F_1(\text{P})$
WORDS	.923	.350
WORDS POS	.913	.399
WORDS TK	.920	.553
WORDS POS TK	.919	.583

Table: Pause prediction (**Pause/NoPause**), improved with syntactic features (POS tags, Tree Kernels)

Some Research Questions

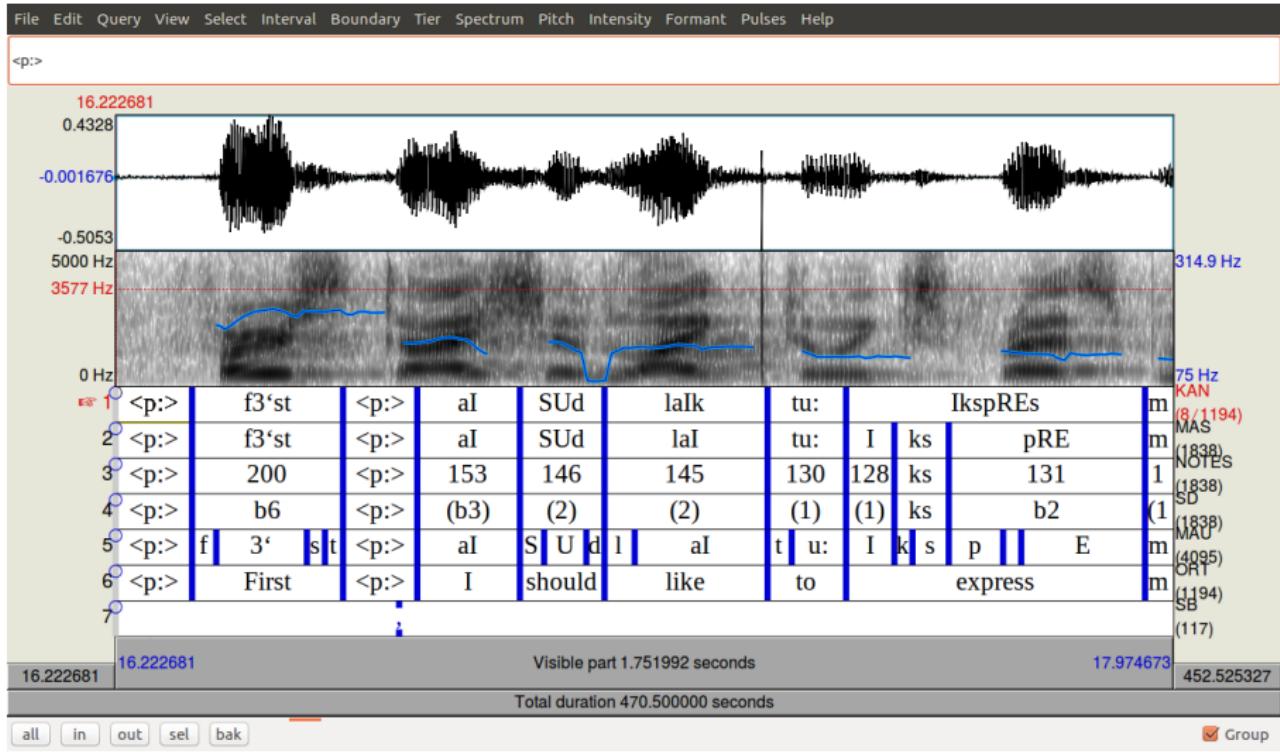
- Evidence for music-speech connection?
- Labeling $\hat{1}$?
- Predict pauses?
- Predict scale degrees?

Audio pipeline

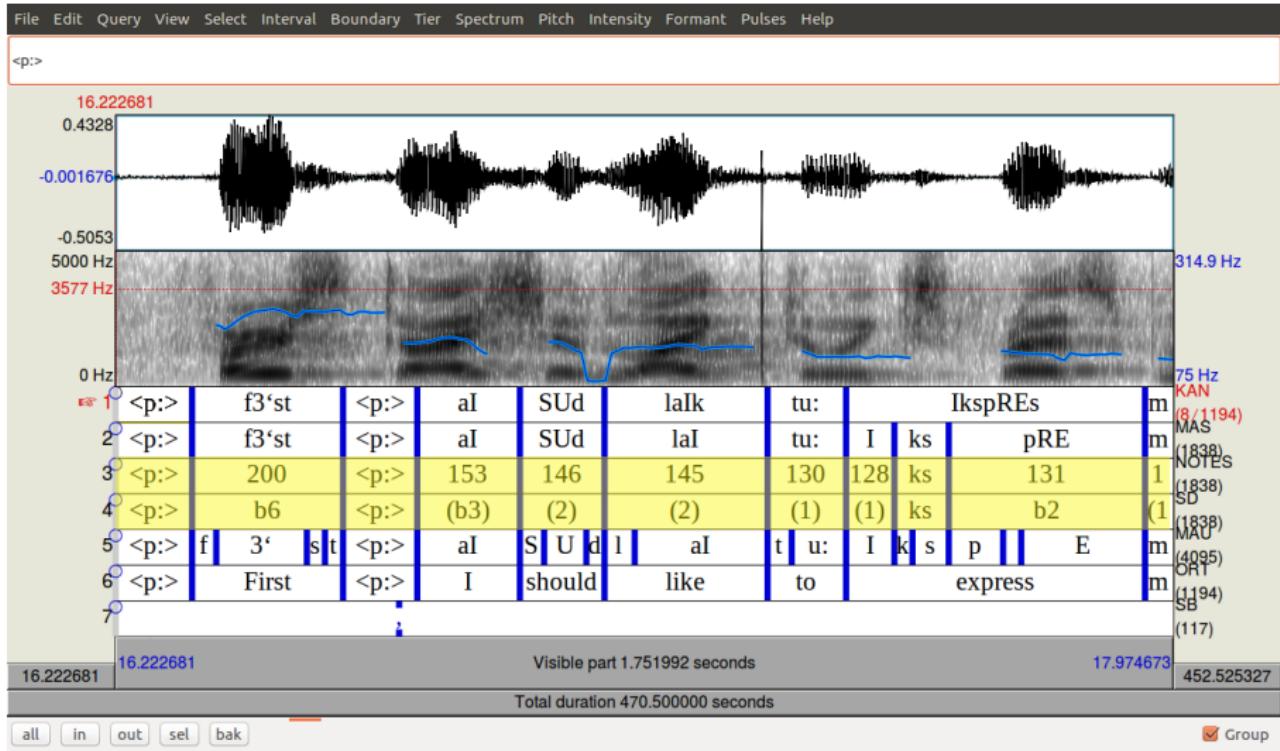


Audio pipeline
for scale degree extraction

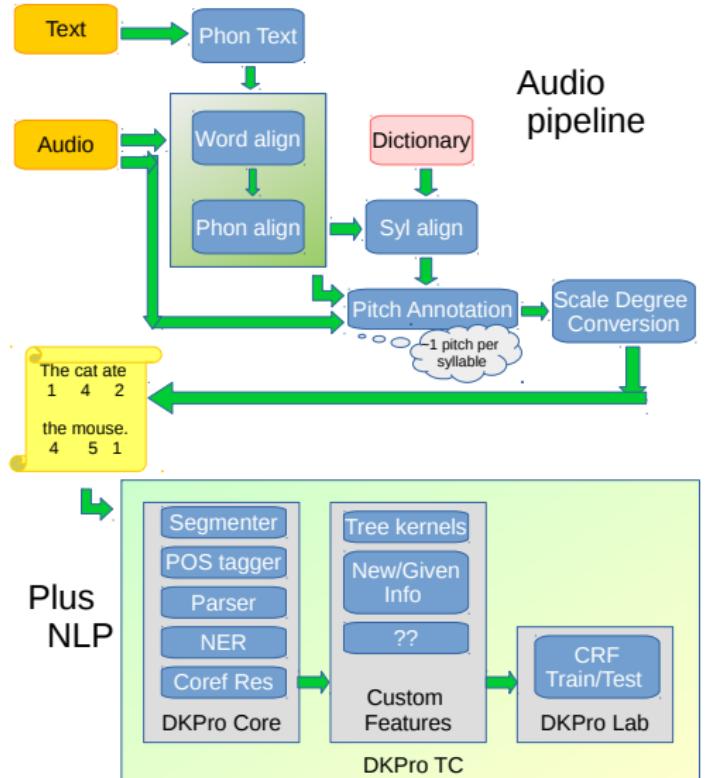
Audio pipeline



Audio pipeline



Entire pipeline



Example: How NLP helps

Google Voice: No *focus/contrastive topic* modeling.

Discussion: Who ate what at the party



Who ate the potato chips? **Marty** ate the potato chips.



What did Marty eat? Marty ate the **potato chips**.



Which chips did Marty eat? Marty ate the **potato** chips.

Some Research Questions

- Evidence for music-speech connection? *Yes.*
- Labeling $\hat{1}$? *Yes.*
- Predict pauses? *POS tags and syntactic tree kernels give 20 pp improvement!*
- Predict scale degrees? *In progress...*

MusLing: Next Steps

- Predicting scale degrees: evaluation

MusLing: Next Steps

- Predicting scale degrees: evaluation
- Predicting scale degree: extracting useful linguistic features

MusLing: Next Steps

- Predicting scale degrees: evaluation
- Predicting scale degree: extracting useful linguistic features
- MaryTTS speech synthesis: implementation and evaluation with humans

MusLing: Next Steps

- Predicting scale degrees: evaluation
- Predicting scale degree: extracting useful linguistic features
- MaryTTS speech synthesis: implementation and evaluation with humans
- ???

The MusLing Project

The End

Thank you for your attention