ELECTRICAL & COMPUTER
ENGINEERING

TiaL
Transformation, Interpretation
& Analysis of Language

Prosody Basics

ECE 596D/LING 580G – Conversational AI

Trang Tran
University of Washington

Agenda

- Announcements:
  - Final presentations + demo (15 mins); "poster" session
  - Monday, June 10, ECE 303, 2-4pm
  - Amazon guests
- Background
  - Prosody: definitions & conventions
  - Prosody in human communication
  - Prosody in language technology
- Prosody Control in Alexa
  - Quick test interface
  - Speech Synthesis Mark-up Language (SSML)
- Project work time

Outline

- **Background**
    - Prosody: definitions & conventions
    - Prosody in human communication
    - Prosody in language technology
- Prosody Control in Alexa
    - Quick test interface
    - Speech Synthesis Mark-up Language (SSML)
- Project work time

Background: Prosody

- Aspects of speech communicating information beyond written words
  - PERmit vs. perMIT; RECord vs. reCORD (meaning)
  - "Mary knows many languages, you know." vs.

    "Mary knows many languages *(that)* you know." (syntax)
  - "You want coffee?" vs. "You want coffee." (intent)
  - "Yeah, sure." vs. "YEAH! SURE!" (sentiment)
- Prosody in human communication: common & essential
- Prosody in AI systems: important but limited
  - Speech (input) understanding: recognition, parsing
  - Speech (output) generation: mostly neutral

Prosody Representation

- Symbolic level:
  - Prominence: relative salience of elements in utterance
  - Phrasing: grouping of words in utterance
- Acoustic cues:
  - Timing, duration
  - Pitch (F0), intonation patterns
  - Energy

➔Acoustic cues individually and in combination signal prominence and phrasing

- Correlates:
  - Increased pitch range, loudness for emphasis
  - Pauses, longer durations preceding phrase boundaries

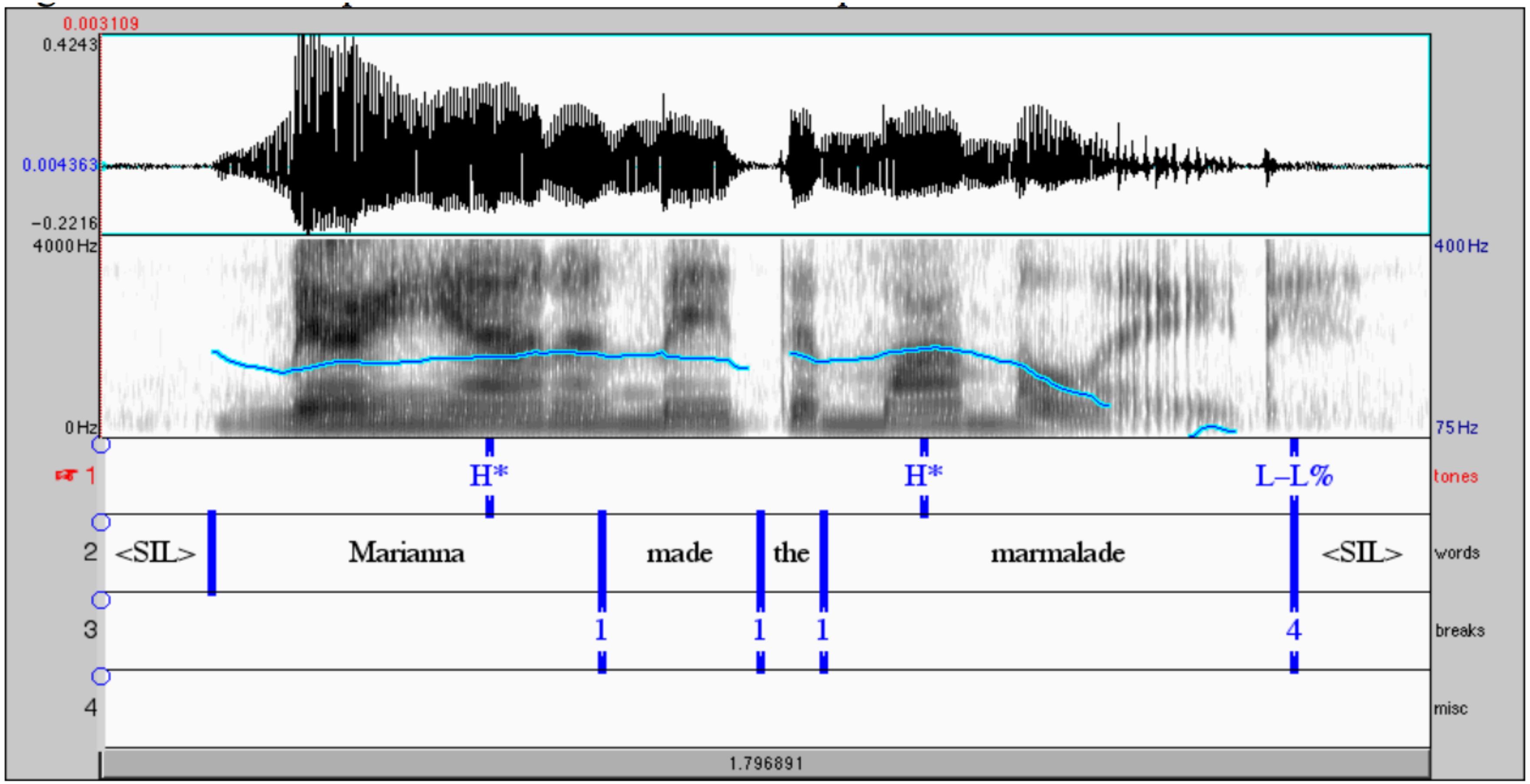
➔ Mapping between acoustic & symbolic levels is complex; challenging to annotate

ToBI Example

Common annotation system: ToBI

Sequence of H(igh) & L(ow) tones

Break indices: 0-4



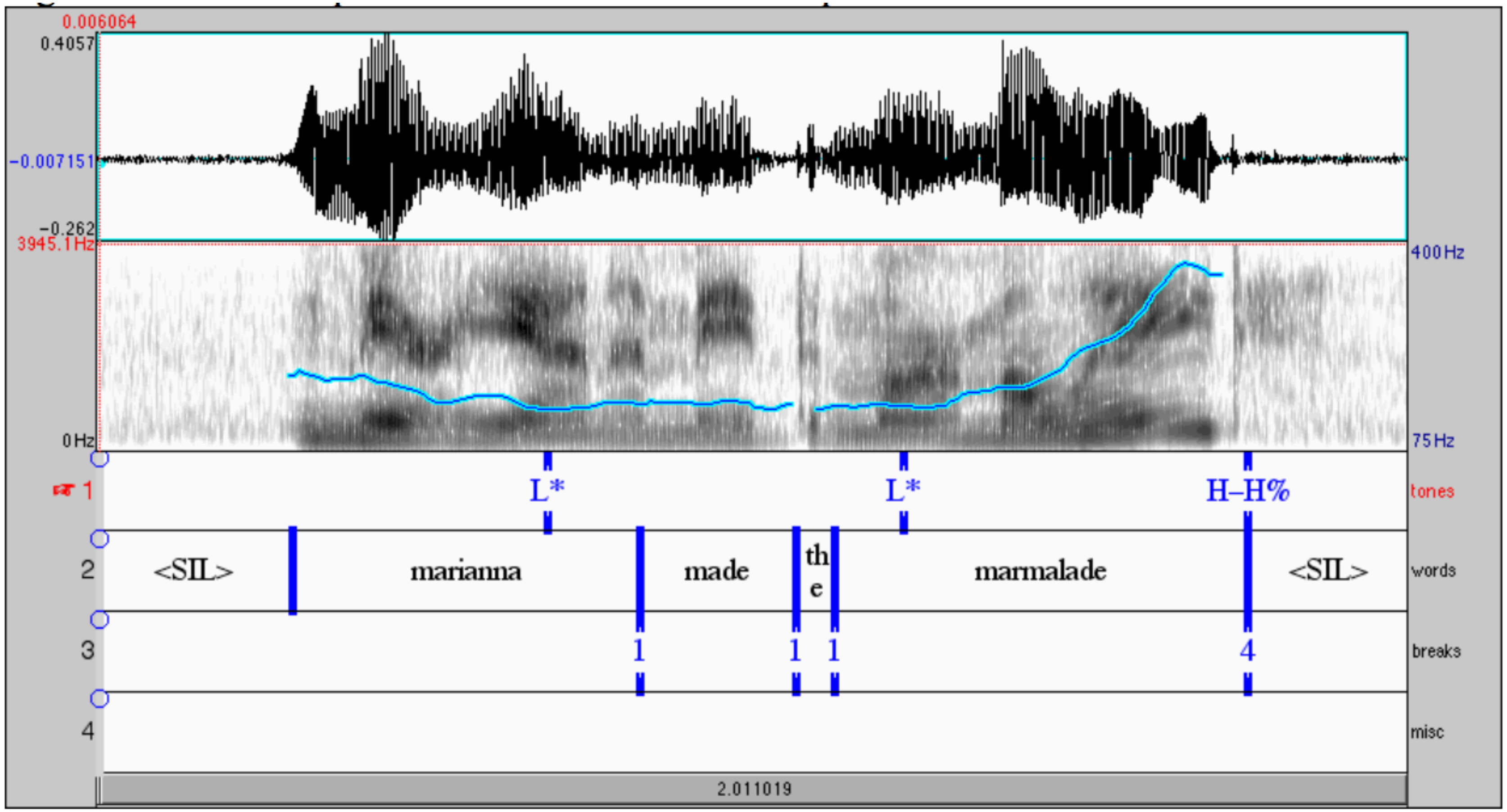
From:

https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/lecture-notes/chapter2_3/

ToBI Example

Common annotation system: ToBI

- Sequence of H(igh) & L(ow) tones
- Break indices: 0-4

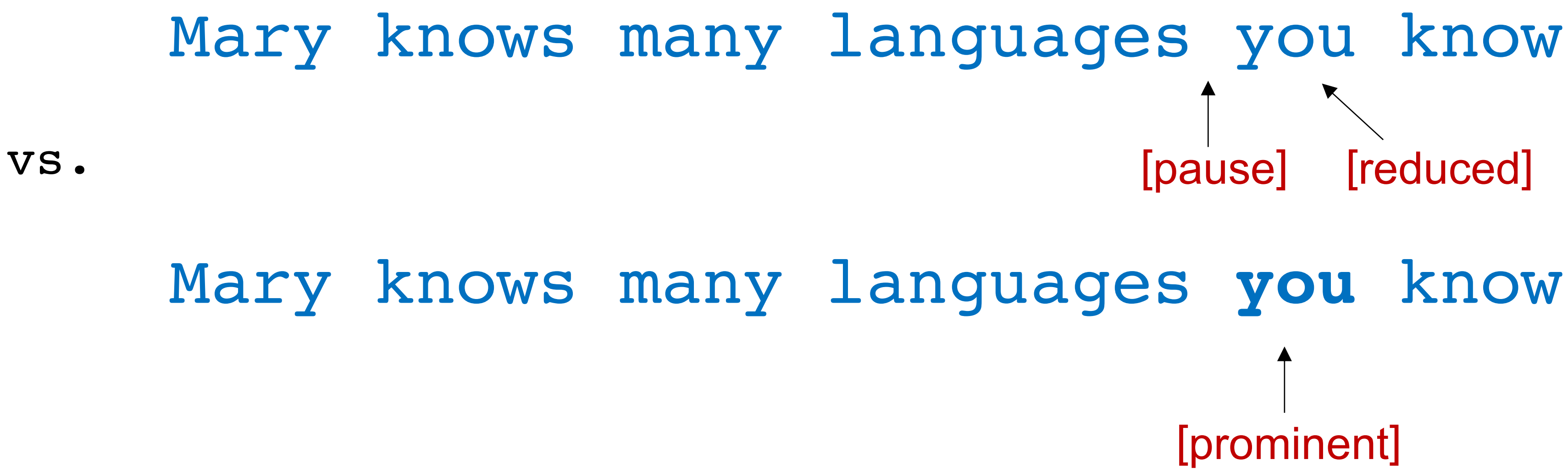


From:

https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/lecture-notes/chapter2_3/

Prosody: Relation to Syntax & Meaning

- Relation to syntax
  - Prosodic boundaries correlate with syntactic boundaries (Grosjean et al., 1979)
  - Resolve structural ambiguities (Price et al., 1991)



Mary knows many languages you know

[pause] [reduced]

vs.

Mary knows many languages **you** know

[prominent]

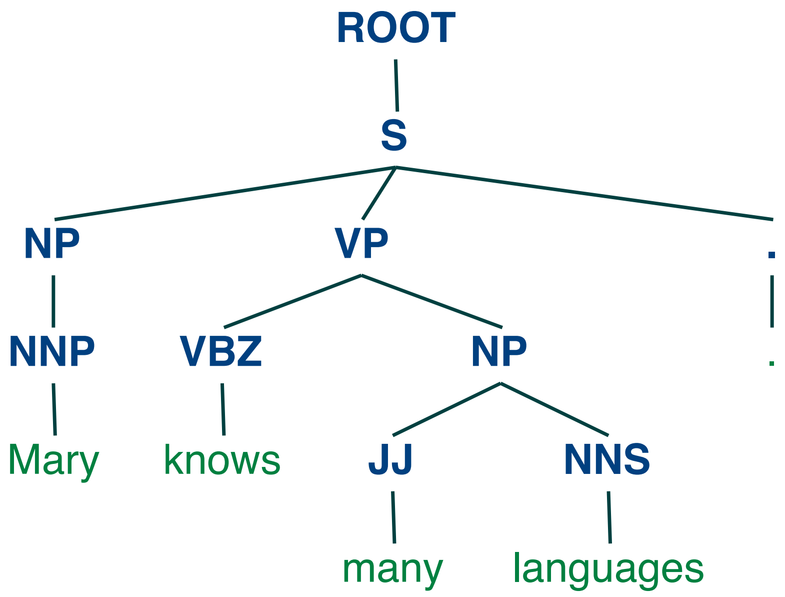
Prosody in Parsing

- Parsing: Identifying syntactic structure of a sentence
- Challenges for speech data:
  - Lacks common cues in written text
  - Disfluencies: filled pauses, [edits] repairs
- Previous works:
  - Gain from prosody was negative or minimal
  - Need explicit (expensive) annotations (ToBI)

Input:

Mary knows many languages.

Output:



Input with disfluencies:

[she knew] mary knows many uh languages

Prosody: Relation to Syntax & Meaning

- Relation to syntax
  - Prosodic boundaries correlate with syntactic boundaries (Grosjean et al., 1979)
  - Resolve structural ambiguities (Price et al., 1991)

- Relation to meaning
  - Prominence signals entity importance (Grosz, 1977)
  - Prominence signals given/new information (Halliday, 1967; Huang & Hirschberg, 2015)

Mary knows many languages

vs.

Mary knows many **languages**

Prosody: Relation to Syntax & Meaning

- Relation to syntax
  - Prosodic boundaries correlate with syntactic boundaries (Grosjean et al., 1979)
  - Resolve structural ambiguities (Price et al., 1991)

Useful for understanding structure (parsing)

- Relation to meaning
  - Prominence signals entity importance (Grosz, 1977)
  - Prominence signals given/new information (Halliday, 1967; Huang & Hirschberg, 2015)

Useful for generation (concept-to-speech)

Prosody in Generation

- TTS (text-to-speech):
  - input = unconstrained text
  - controlling prosody:
    - text analysis
    - prosody (ToBI) prediction
    - waveform generation/modification

context independent

- CTS (concept-to-speech):
  - input = intent-defined text
  - controlling prosody:
    - from intent
    - waveform generation/modification

predefined schemata

intensive signal processing; prone to distortion

- External prosody control:
  - Markup languages: **SSML**, Sable

available in most commercial systems

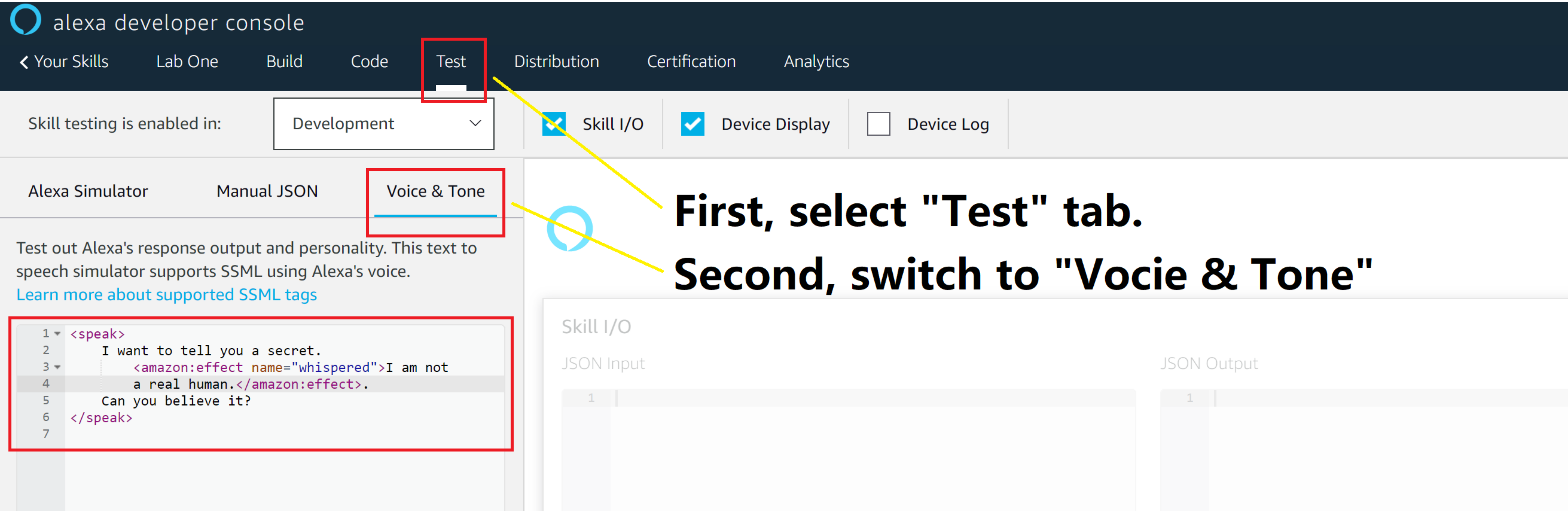
Common Challenges

- Systems like ToBI
    - expensive to annotate
    - even experts disagree
    - language-dependent
- Integration of discrete (words) with continuous (acoustics) signals
- Studies on prosody: mostly in controlled, read speech
- In many tasks: ultimate goal, reference signal is still tied to words
    - Recognition, parsing
    - TTS, CTS: good quality on neutral, read style

Outline

- Background
  - Prosody: definitions & conventions
  - Prosody in human communication
  - Prosody in language technology
- **Prosody Control in Alexa**
  - Quick test interface
  - Speech Synthesis Mark-up Language (SSML)
- Project work time

Quick Test Interface



Skill testing is enabled in: Development

☑ Skill I/O ☑ Device Display □ Device Log

Test out Alexa's response output and personality. This text to speech simulator supports SSML using Alexa's voice.
Learn more about supported SSML tags

```
<speak>
    I want to tell you a secret.
        <amazon:effect name="whispered">I am not
        a real human.</amazon:effect>.
    Can you believe it?
</speak>
```

First, select "Test" tab.

Second, switch to "Vocie & Tone"

SSML

- Speech Synthesis Markup Language
  - Giving users (limited) control over prosody – can change pitch, speech rate, voice, etc.
  - https://developer.amazon.com/docs/custom-skills/speech-synthesis-markup-language-ssml-reference.html
  - https://developer.amazon.com/docs/custom-skills/speechcon-reference-interjections-english-us.html
- Demo

Outline

- Background
  - Prosody: definitions & conventions
  - Prosody in human communication
  - Prosody in language technology
- Prosody Control in Alexa
  - Quick test interface
  - Speech Synthesis Mark-up Language (SSML)
- **Project work time**

Extra Slides

Prosody in Education Applications

- Assessment
  - Prosodic & rhythm sensitivity correlates with reading ability
  - Better readers produce pitch & pause patterns that align with syntax
- Implications
  - Early exposure to diverse prosody affects later academic success
  - Interactive learning environments are critical, but not always available in low socio-economic communities
- Social robots
  - Adaptive robots encourage learning, especially with expressive prosody
  - https://youtu.be/4zuaL7hIYq0