MONASH University

Department of Electrical and Computer Systems Engineering

ECE4095 Final Year Project, Semester 1, 2015

Ruben Bloom

Paralinguistic Speech Analysis

Supervisor: Professor Tom Drummond (formerly Dr. Wai Ho Li)

1 Project Aim

1. To identify paralinguistic speech features which contribute to a speaker's *speaking style,* i.e. *how they sound.*
2. To automatically extract these features from recorded speech.
3. To use the developed tools to analyse a test set of speech recordings. This should verify the practical usefulness of the tools and provide insight into the differences between speakers and groups of speakers.

The following features were extracted and analysed: pauses, utterances, pitch statistics, and finality patterns. See below for definitions.

2 What are paralinguistics?

While *linguistics* are what you say, *paralinguistics* are *how you say it.* Paralinguistic speech features are all those aspects that go beyond the words themselves, such as pitch, variations in pitch, pauses, length of utterances, speech rate, umms and ahhs, and more.

Our paralinguistics greatly shape how others perceive us in presentations, interviews, and everyday conversations. They're what separate the boring speakers from the dynamic, exciting, and persuasive! They matter, and this project is about using technology to study them.

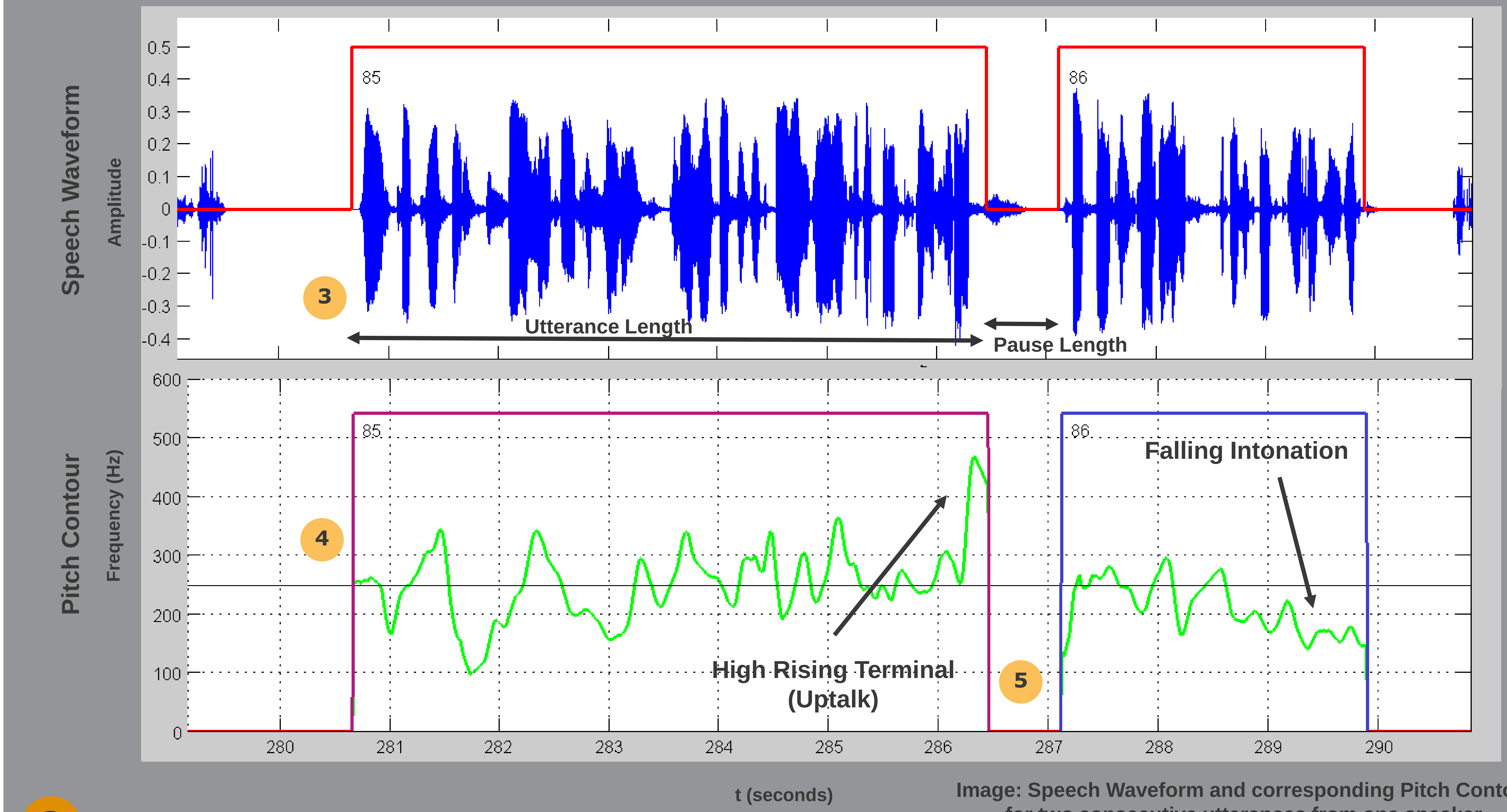

Image: Speech Waveform and corresponding Pitch Contour for two consecutive utterances from one speaker.

3

A Voice Activity Detection (VAD) algorithm identifies the presence and absence of speech in a recording. The output is used to segment the recording into separate utterances and to identify pauses. Additionally, pause and utterance duration statistics are one element of speaking style.

4

The RAPT pitch tracking algorithm is used to calculate the pitch contour from speech recordings. Global pitch statistics such as mean and variance are included in the speaking style profile.

5

The segmentation of utterances is combined with pitch tracking to detect Finality Patterns: the movement of pitch at the end of utterances.

In spoken English, declarative sentences typically end with a drop in pitch (falling intonation), whereas questions end with a rise. However, Australians are known for increasing the pitch even at the end of statement, making them sound like questions. This is known as *High Rising Terminal* (HRT) or "uptalk". Without judging whether uptalk is good or bad, the system detects it.

Example Speaking Style Profiles*

Name	Emily	Jess
Mean f0	249 Hz	167Hz
f0 SD	67 Hz	44Hz
Speech/Pause %	79/21	78/22
Mean Pause Duration	660ms	775ms
Uptalk %	11%	42%
Falling Intonation %	34%	17%

*Real speakers with names changed.

6 SPEECH ANALYSIS RESULTS

4-5 minute speech recordings were analysed from two groups:

- TED talk presentations, where each video had received 1M or more views. They are presumed to be highly charismatic (n =7).
- Oral Presentations from 2nd Undergraduate Psychology Students (n = 7).

TED presenters had higher pitch and pitch variation on average, their pauses were longer, and they paused more of the time.

Student presenters ended 16% of their utterances with High Rising Terminal on average, vs 6% for TED presenters. Students ended 10% of their sentences with Falling Intonation, vs. 26% for TED presenters.

These automatically extracted results match those expected for the difference between experienced, professional speakers compared with the inexperienced: more dynamicness through variation of pitch, well timed pauses, and strong emphatic speech.