
meamh
야옹
meong
mjav
miaou
mjau
ونونون
мяу
ņau
miauw
meow
mnau
mjá
miyav
miao
mjay
ニャー
喵


?

Automatic Language Identification

(LiD)

Alexander Hosford

@awhosford

The LiD Problem

- The identification of a given spoken language from a speech signal
- 94% of global population speak only 6% of world languages

Real World Situations

- Automation of LiD is desirable
- Offers many benefits to international service industries
- Hotels, Airports, Global Call Centres

Language Differences

- Languages contain information that makes one discernable from the other;
  - Phonemes
  - Prosody
  - Phonotactics
  - Syntax



Human Abilities

- Bias towards native language arises in infancy
- Prosodic features some of the first cues to be recognised
- Humans can make a reasonable estimate on language heard within 3-4 seconds of audition
- Even unfamiliar languages may be plausibly judged

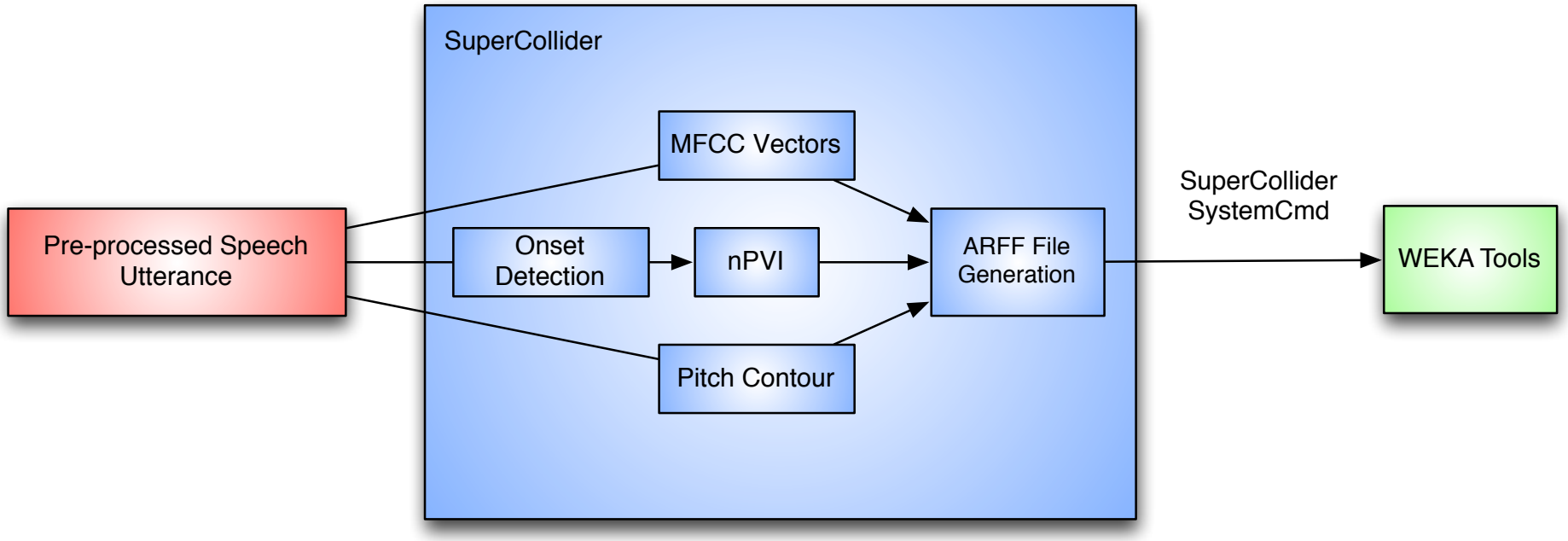
Previous Attempts

- Attempts as early as 1974 – USAF work, therefore classified.
- Methodological Investigations as early as 1977 (House & Neuburg)
- Studies for the most part center on phonotactic constrains and phoneme modeling
- Raw acoustic waveforms have been visited (Kwasny 1993)

A Simpler Approach?

- Phonotactic approaches require expert linguistic knowledge
- Phoneme and phonotactic modeling time consuming
- Given the speed of human LiD abilities, discrimination most likely based on acoustic features

System Overview



Feature Extraction

- Spectral Information – MFCC Vectors
- Pitch contour information
- Handled by SCMIR Library (Collins, 2010)
- Speech Rhythm – Normalised Pairwise Variability Index (nPVI)

Feature Type	Implemented Measure
Spectral Content	MFCC Vector uGen
Pitch Contour	Tartini uGen
Speech Rhythm	nPVI function

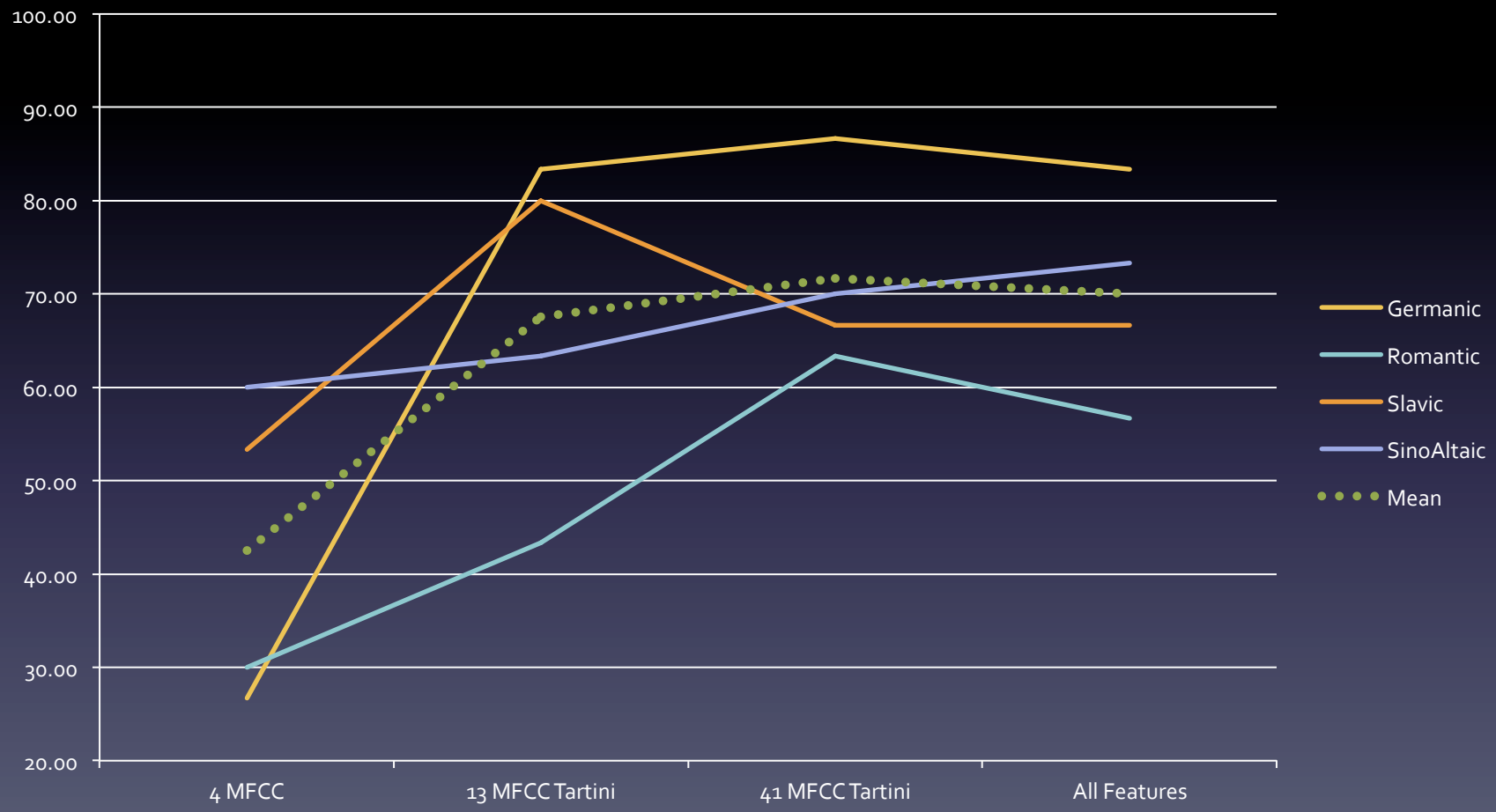
Classification

- Handled by the WEKA toolkit
- Built in Multilayer Perceptron
- Called from the command line through a SuperCollider system command

Comparisons Made

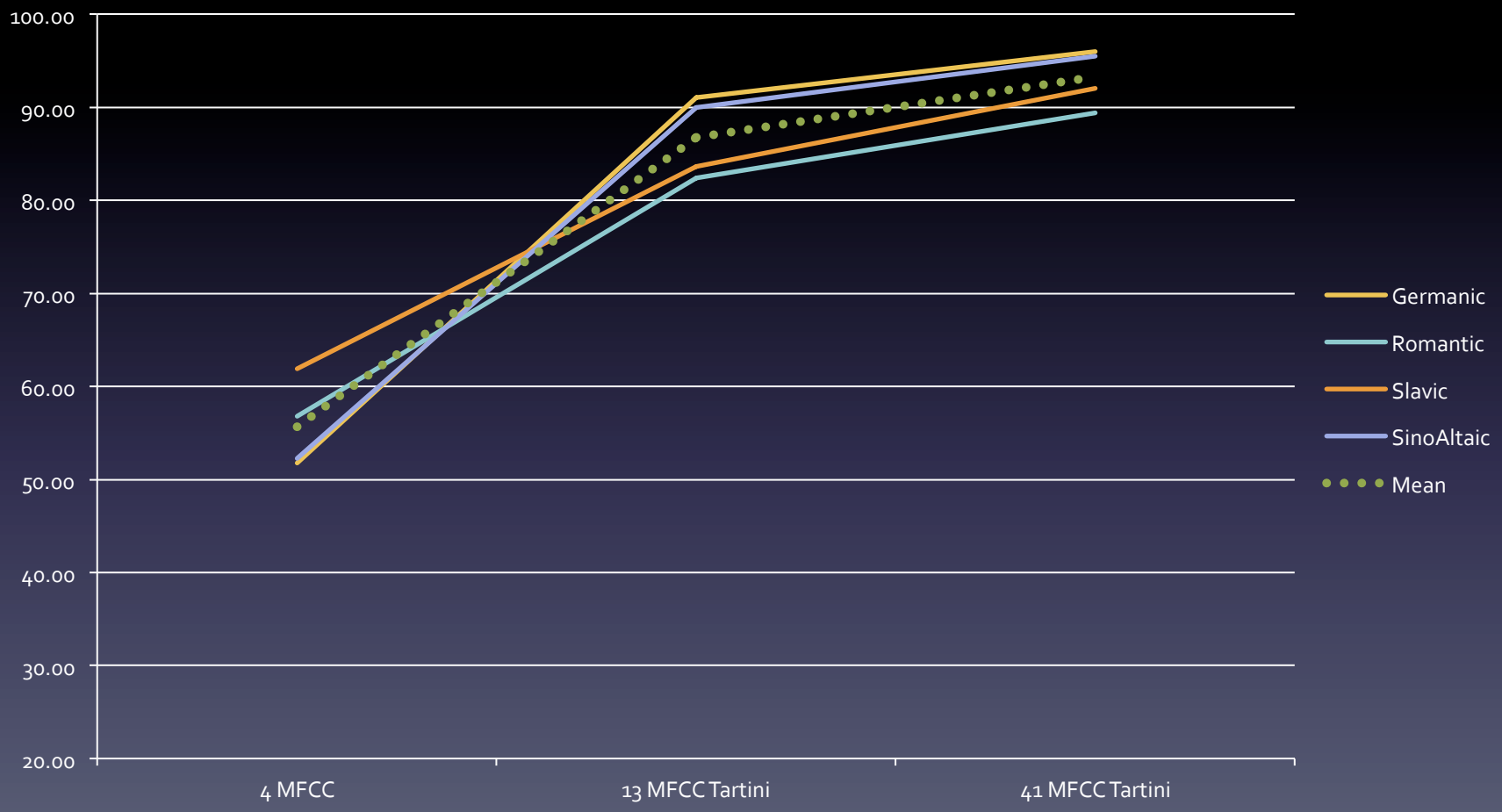
- 66 language pairs from 12 languages
- A comparison within language families
- A comparison of all 12 languages
- Averaged & segmented data

Results Within Families



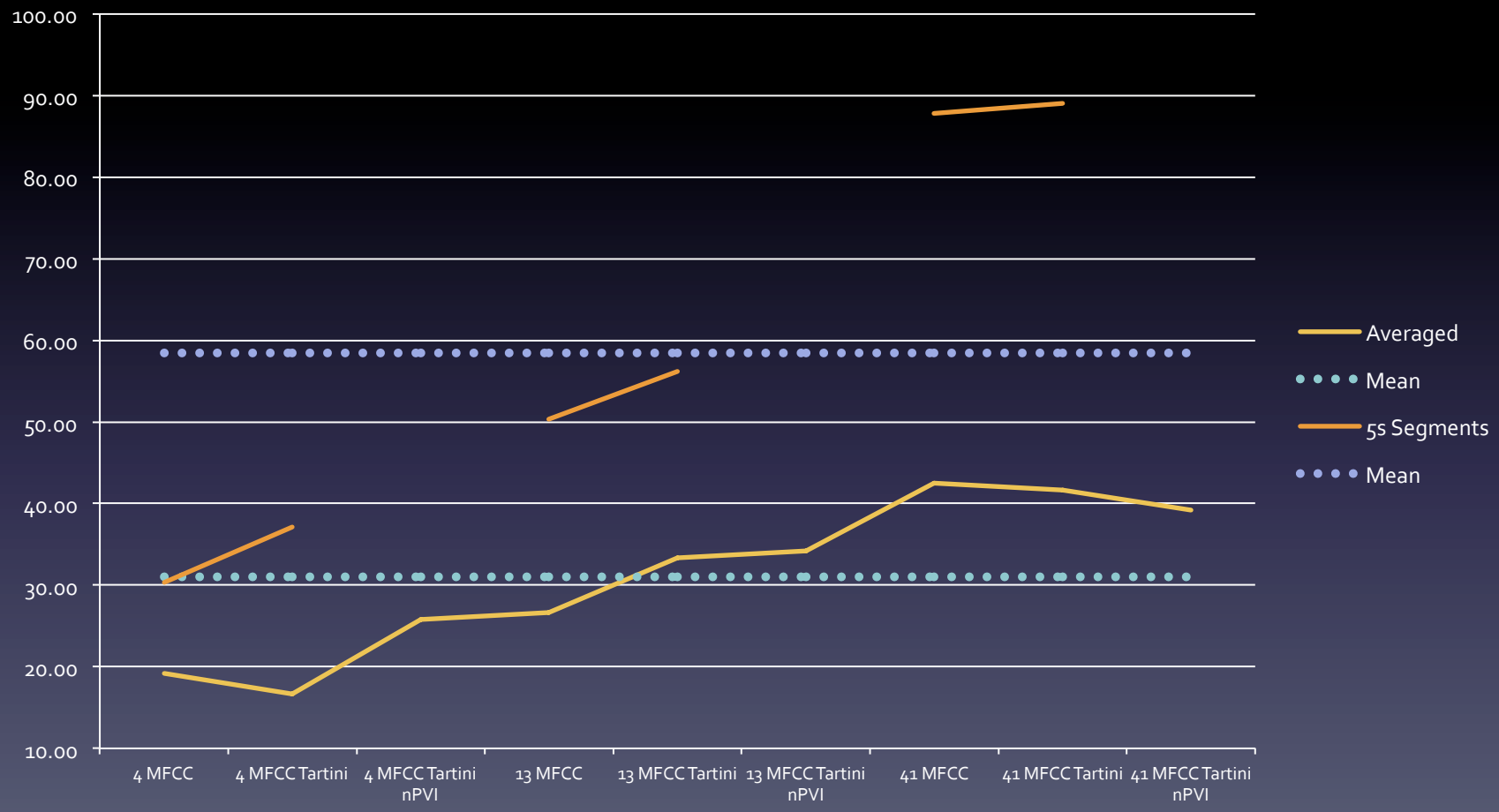
*Data averaged across files

Results Within Families



*Data in 5 second segments

Results From All Languages



Extensions

- nPVI function to use vowel onsets
- Phonemic Segmentation
- A Larger dataset
- Better Efficiency
- Real-time operation

Robustness

- Real world signals are very different from processed ‘clean’ data.
- ‘Ideal’ LiD systems – independent & robust
- A need to analyse only the part of the signal that matters.

CASA

- A computational modeling of human 'Auditory Scene Analysis' (Bregman 1990)
- Separation of signal into component parts and reconstitution into meaningful 'streams'

Using Noisy Signals

I’m open to suggestions!

alexanderhosford@googlemail.com
07814 692 549
@awhosford