Investigating the forensic applications of global and local temporal representations of speech for dialect discrimination

Leah Bradshaw, Vincent Hughes, and Eleanor Chodroff

Department of Language and Linguistic Science
University of York
INTRODUCTION

Forensic phonetics

Voice analysis

Voice comparison
Speaker classification

Process of determining speaker-specific features (e.g., gender, age, dialect, idiosyncratic speech markers, etc.) using:

- Auditory analysis
- Acoustic-phonetic analysis
- Automatic speaker recognition approaches
Acoustic-phonetic analysis frequently involves court-presentable measurements that are strongly focused on segmental information:

- Formants
- F0
- Voice onset time

But what about suprasegmental information, and specifically information about a speaker’s rhythmic pattern?
INTRODUCTION

Rhythm in speaker classification

Previous studies demonstrate some utility of rhythm for dialect discrimination and forensic purposes

Limited in its application in research and casework

Rhythm depends on some temporal representation of speech

Rhythm: Temporal characteristics of a spoken utterance

How can temporal characteristics of a spoken utterance be represented in an acoustic-phonetic analysis? In an ASR analysis?
Global temporal representations

Long-term alternations in vocalic and consonant intervals which may approximate the rhythmic pattern of speech

**Rhythm Metrics:** measures examining the degree of variability in the duration of pre-specified intervals (e.g., vowels, consonants, CV sequences, adjacent intervals, etc.)

Rasmus et al., 1990, Grabe and Low 2002, Dellwo 2006
Rhythm in speaker classification

Syllable vs stress-timed distinctions

**Syllable-timed:** equal syllable durations

**Stress-timed:** equal stressed syllable durations (more variability between stressed and unstressed syllables)

Problematic: too coarse – but, possibly a place to start

Local temporal representations

Delta (Δ) and delta-delta (ΔΔ) features: Reflect the change in spectral properties between adjacent temporal frames and the acceleration of that change

Common in ASR systems

e.g., Lee et al. 1990, Matsui and Furui 1990, Gish and Schmidt 1994
GOALS

1) Analyze rhythmic profile of four varieties of British English: Cambridge, Multicultural London English, Leicester, and Punjabi-Leicester

2) Investigate the utility of global RMs for discriminating among the dialects

3) Compare global and local temporal representations for dialect discrimination
Introduction

Corpus Description

Global: Rhythm Metrics

Local: Deltas and Delta-deltas

Discussion
### Four British English Dialects

<table>
<thead>
<tr>
<th>Non-contact (Anglo)</th>
<th>Contact (Ethnic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambridge English (CE)</td>
<td>Multicultural London English (MLE)</td>
</tr>
<tr>
<td>Leicester English (LE)</td>
<td>Caribbean descent</td>
</tr>
</tbody>
</table>

- Leicester ("Midlands")
- "South"

**International Varieties of English (IViE) corpus:** 12 CE, 12 MLE, age 16

**Wormald (2016):** 8 LE, 22 PLE, ages 20–53
Introduction

Corpus Description

Global: Rhythm Metrics

Local: Deltas and Delta-deltas

Discussion
GLOBAL MEASURES:

Rhythm Metrics

- **stdevV**: Standard deviation of vocalic interval duration
- **stdevC**: Standard deviation of consonantal interval duration
- **VarcoV**: Coefficient of variation for the vocalic interval duration
GLOBAL MEASURES:
Rhythm Metrics

nPVI-V  Pairwise Variability Index for vocalic interval durations
nPVI-C  Pairwise Variability Index for consonant interval durations
nPVI-CV Normalised pairwise variability index for summed consonantal and vocalic interval durations
Cambridge, MLE: Praat EasyAlign for British English

Leicester varieties: Alignments accompanied the recordings

All phone alignments were manually adjusted

Consonantal and vowel intervals determined based on the phone alignments

RM5s measured with the Duration Analyzer Praat script
RESULTS:

Rhythm Metrics

Dialect significantly improved model fit

No gender differences
RESULTS:

Rhythm Metrics

Cambridge English: higher stdev-V, VarcoV, nPVI-CV

MLE: average

Leicester English: higher VarcoV

Punjabi Leicester: lower stdev-V, VarcoV

All relative to the average production across all four dialects
RESULTS:

Rhythm Metrics

Cambridge English: lower stdv-C

MLE: lower stdv-C, nPVI-V, nPVI-C

Leicester English: higher stdv-C, nPVI-V, nPVI-CV

Punjabi-Leicester: higher stdv-C
lower nPVI-V

All relative to the average production across all four dialects
RESULTS:

Rhythm Metrics

~ Midlands vs. South

~ Anglo vs Contact

Purity: 0.64
Outline

Introduction

Corpus Description

Global: Rhythm Metrics

Local: Deltas and Delta-deltas

Discussion
METHODS:

Δ and ΔΔs

MFCCs

Voice activity automatically detected + manually corrected

20 ms frames, shifted by 10 ms

0–4000 Hz

CMVN applied for room/equipment normalization
"Δs and ΔΔs

Deltas: change between MFCCs in adjacent frames

Delta-deltas: change between deltas in adjacent vectors

Averaged for each recording

MFCCs not included in the analysis"
RESULTS:
Deltas and delta-deltas

Cluster plot
Purity: 0.44
Significant differences in RMs among four British English dialects

CE and LE more stress-timed—but in different ways

MLE and PLE more syllable-timed—but in different ways

Combination of RMs can be used as a Rhythmic Profile
Rhythmic profile is a useful feature in dialect discrimination

**Issue:** RMs somewhat correlated

**Future directions:** Which RMs and combinations of RMs are indeed best and least redundant?

Examine whether these results hold for dialects collected in a single corpus
Proof of concept: Global temporal representations > local temporal representations for dialect discrimination

Demonstrates need for global temporal representation in automatic speaker and language recognition systems
(some work done already)

Forensic application of RMs: directly interpretable, court presentable

Adami et al. 2003, Shriberg et al. 2005, Dehak et al. 2007
Thank you!

Thanks to:
Jess Wormald
Paul Foulkes
Peter French
Sam Hellmuth