# CONTRASTING EMOTION-BEARING LAUGHTER TYPES IN MULTIPARTICIPANT VOCAL ACTIVITY DETECTION FOR MEETINGS

Kornel Laskowski
Language Technologies Institute, Carnegie Mellon University, Pittsburgh PA, USA
Cognitive Systems Lab, Universität Karlsruhe, Karlsruhe, Germany

## Goal

Improve laughter detection for upstream conversation-processing tasks.

HOW?

Partition laughter into types and focus on the detection of that type which is most useful to an upstream task, e.g.

$$\underset{\text{all laughter}}{\mathcal{L}} \equiv \underset{\text{voiced laughter}}{\mathcal{L}_V} \cup \underset{\text{unvoiced laughter}}{\mathcal{L}_U}$$

## Is Voiced Laughter More Useful Than All Laughter?

**Candidate upstream task**:

Detect regions of **involved** speech, given a vocal activity segmentation.

**Results (accuracies, %)**:

| Segmentation | Accuracy, % | | |
|---|---|---|---|
| | TrainSet | DevSet | TestSet |
| guess, majority | 73.7 | 72.9 | 73.7 |
| $\mathcal{S}$ | 72.7 | 74.8 | 75.2 |
| $\mathcal{L}$ | 79.2 | 80.4 | 80.6 |
| $\mathcal{L} \cap \mathcal{S}$ | 84.3 | 82.7 | 83.0 |
| $\mathcal{L}_V$ | 80.2 | 81.2 | 81.4 |
| $\mathcal{L}_V \cap \mathcal{S}$ | **84.4** | **82.9** | **85.6** |
| $\mathcal{L}_U$ | 77.7 | 76.4 | 77.4 |

## Data

1. ICSI Meeting Corpus (Janin et al, 2003)
2. 75 meetings, over 66 hrs of multichannel audio
3. reference segmentation of speech (Shriberg et al, 2004)
4. reference segmentation of laughter (Laskowski & Burger, 2007)
5. reference segmentation of involved speech (Wrede et al, 2005)

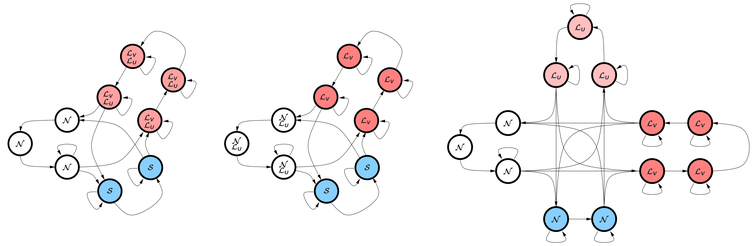
## Acoustic Detection Experiments

**Acoustic models**:

- close-talk microphone channels
- standard MFCC features (39)
- crosstalk NLED features (2)
- Gaussian mixture models (64)

**Topologies**:

A1 B1 C1



- 100 ms frame step (and frame size)
- A0, B0, and C0 are the ergodic counterparts
- B2, and C3 contain optimal duration constraints
- **joint participant** topologies are Cartesian products

## Findings

1. Ignoring $\mathcal{L}_U$ improves the upstream detection of hotspots.
2. Modeling $\mathcal{L}_U$ with silence improves the acoustic detection of $\mathcal{L}_V$.
3. Modeling $\mathcal{L}_U$ explicitly improves the acoustic detection of $\mathcal{L}_V$.
4. For ergodic HMM topologies, joint participant decoding improves the acoustic detection of $\mathcal{L}_V$.
5. For non-ergodic HMM topologies, independently decoding participants improves the acoustic detection of $\mathcal{L}_V$.

## Acoustic Detection Experiments

**Results (*F*-scores)**:

| System | Indep. Participant | | | Joint Participant | | |
|---|---|---|---|---|---|---|
| | $\mathcal{S} \cup \mathcal{L}$ | $\mathcal{L}$ | $\mathcal{L}_V$ | $\mathcal{S} \cup \mathcal{L}$ | $\mathcal{L}$ | $\mathcal{L}_V$ |
| A0 | 75.4 | 30.9 | — | 78.1 | 31.7 | — |
| A1 | 76.3 | 32.6 | — | 79.5 | 34.5 | — |
| B0 | 78.3 | — | 34.6 | 79.5 | — | 34.2 |
| B1 | 79.0 | — | 36.4 | 80.9 | — | 37.3 |
| B2 | 81.7 | — | 46.0 | *intractable* | | |
| C0 | 71.6 | 25.9 | 32.9 | 76.0 | 26.2 | 27.3 |
| C1 | 72.6 | 27.6 | 34.4 | 78.9 | 30.4 | 31.2 |
| C3 | 74.6 | 32.4 | **47.7** | *intractable* | | |

Notes:

1. $\mathcal{S} \cup \mathcal{L}$ is all vocalization (versus silence).
2. It is intractable to impose large duration constraints in joint participant topologies.


kornel@cs.cmu.edu