# A **reliability-aware** model

# for

# intelligibility classification

# in

# pathological speech


**Naveen Kumar, Shrikanth Narayanan**

*Signal Analysis and Interpretation Lab*

*University of Southern California*

# Overview

1. Pathological Speech
2. Defining reliability
3. Reliability-aware classification model
4. Results
5. Summary

# Pathological Speech

- Atypicality resulting from disease or surgery of the vocal tract
- Reduced speech intelligibility
- Decrease of intelligibility might be perceived by different factors

# Database

## NKI CCRT Speech Corpus

- Speech Intelligibility before and after treatment.
- Released during Interspeech 2012 speaker trait challenge[1]
- 2385 sentence level utterances
- Ratings thresholded to
  - intelligible (I)
  - non-intelligible (NI)



[1] Schuller, Björn, et al. "The INTERSPEECH 2012 Speaker Trait Challenge." *INTERSPEECH*. 2012.

# Feature Subsets

Utterance level feature subsystems adapted from [2]

- **Prosody** (6)
  - Pitch - L0 norm, polynomial fit, variance
- **Pronunciation** (2)
  - CMN 39 dim. MFCC, phone duration
- **Voice Quality** (5)
  - HNR, Jitter, Shimmer

[2] Kim, Jangwon, et al., "*Automatic intelligibility classification of sentence level pathological speech",* in Computer Speech and Language, 2013

# Defining Reliability



How was the label **Y** assigned given features **X?** **(discriminative)**

**reliable** (R=1)

**unreliable** (R=0)

# Reliability of labels

How was the label **Y** assigned given features **X?** **(discriminative)**

**reliable** (R=1)

**unreliable** (R=0)

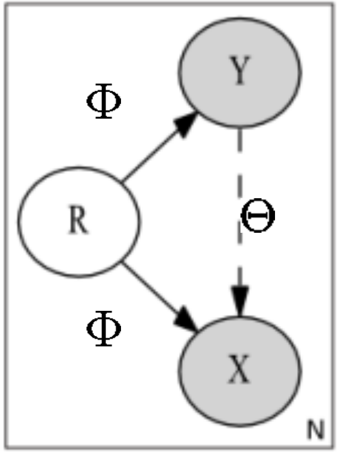
**Why model label reliability?**

- human annotations are inherently subjective
- noisy features on some samples

# Reliability formulation



$$Pr(X,Y|R) = Pr(X,Y;\Theta)^R[Pr(X;\Phi)Pr(Y;\Phi)]^{1-R}$$




label

feature

- R={0,1} : reliable at random model
- $\Theta$ : data dependent reliable model
- $\Phi$ : data independent unreliable model

Latent reliability **R** controls **dependence** between data **X** and label **Y**

# Discriminative reliability assumption

- ***unreliable:*** Label **Y** generated independent of data
- ***reliable***: label **Y** generated according to a data-dependent model

$$Pr(Y|X,R) = \underbrace{Pr(Y|X;\Theta)^{R}}_{reliable} \; \underbrace{Pr(Y;\Phi)^{1-R}}_{unreliable}$$

# Discriminative reliability assumption

- ***unreliable:*** Label **Y** generated independent of data
- ***reliable***: label **Y** generated according to a data-dependent model

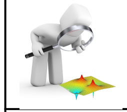
$$Pr(Y|X,R) = \underbrace{Pr(Y|X;\Theta)^R}_{reliable} \underbrace{Pr(Y;\Phi)^{1-R}}_{unreliable}$$

- **data-dependent reliability**

$$Pr(Y,R|X) = Pr(R|X) Pr(Y|X;\Theta)^R Pr(Y;\Phi)^{1-R}$$

- **mixture of experts model**

# Choice of models

| | | | |
|---|---|---|---|
| reliable  $\mathbf{W}$ | ***maximum-entropy (softmax)*** | $Pr(Y_i = k \mid X_i; \boldsymbol{\Theta}) = \frac{e^{W_k{}^T X_i}}{\sum_{j=1}^{K} e^{W_j{}^T X_i}}$ | $\Psi_{ik}(W)$ |
| unreliable  $\boldsymbol{\lambda}$ | ***multinomial*** | $Pr(Y_i = k; \Phi) = \lambda_k$ | $\lambda_k$ |
| reliability model $\mathbf{r}$ | ***logistic regression*** | $Pr(R_i = 1 \mid X_i) = \sigma(\mathbf{r}^T X)$ | $\rho_i$ |

# Parameter estimation

**ML estimation** via EM algorithm

- **E-step** $\gamma_i$

$$Pr(R_i = 1|X_i, Y_i) = \frac{Pr(Y_i|R_i = 1, X_i)\ Pr(R_i = 1|X_i)}{Pr(Y_i|X_i)}$$

- **M step**

  estimate parameters $\mathbf{W}^*$ $\boldsymbol{\lambda}^*$ $\boldsymbol{r}^*$ by maximizing expected data log-likelihood.

# Data log-likelihood

data dependent
reliability model

$$= \sum_{i=1}^{N} \gamma_i \; ln \; \rho_i + (1-\gamma_i) \; ln(1-\rho_i)$$

$\boldsymbol{r}^*$

$\mathbf{W}^*$

reliable and unreliable
models

$$+ \sum_{i=1}^{N} \sum_{k=1}^{K} \underline{Y_{ik}[\gamma_i \; ln \; \Psi_{ik}} + \overline{(1-\gamma_i) ln \; \lambda_k]}$$

$\boldsymbol{\lambda}^*$

# Data log-likelihood

logistic regression
with label $\gamma_i$

$$= \sum_{i=1}^{N} \gamma_i \; ln \; \rho_i + (1 - \gamma_i) \; ln(1 - \rho_i)$$

weighted maxent with
weights $\gamma_i$

$$+ \sum_{i=1}^{N} \sum_{k=1}^{K} \underline{Y_{ik}[\gamma_i \; ln \; \Psi_{ik}} + \overline{(1 - \gamma_i) ln \; \lambda_k}]$$

# Parameter updates and Inference

| | | |
|---|---|---|
| $\boldsymbol{\lambda}_k$ | $\frac{\sum_{i=1}^{N} Y_{ik}(1-\gamma_i)}{\sum_{i=1}^{N} Y_{ik}}$ | one step |
| $\mathbf{r}$ | logistic | gradient ascent (L-BFGS) |
| $\boldsymbol{W}$ | weighted-maxent | gradient ascent (L-BFGS) |

# Parameter updates and Inference

| | | |
|---|---|---|
| $\boldsymbol{\lambda}_k$ | $\frac{\sum_{i=1}^{N} Y_{ik}(1-\gamma_i)}{\sum_{i=1}^{N} Y_{ik}}$ | one step |
| $\mathbf{r}$ | logistic | gradient ascent (L-BFGS) |
| $\boldsymbol{W}$ | weighted-maxent | gradient ascent (L-BFGS) |

**Inference of class labels**

$$Pr(Y=k|Z) = \Psi_k(\mathbf{W}^*, Z)\sigma(\mathbf{r}^{*T} Z) + \boldsymbol{\lambda}_k{}^*[1-\sigma(\mathbf{r}^{*T} Z)]$$

test feature

# Experiment : **Intelligibility classification**

- 5 fold cross validation
- Baseline : Reliability blind classifier, always assumes R=1

| **Feature set** | **Logistic regression** | **Reliability aware** |
|---|---|---|
| voice quality | 58.2 | **59.8** |
| prosody | 67.1 | 66.7 |
| pronunciation | 55.1 | **56.2** |
| feature fusion | 68.0 | 67.8 |

# Analysis : **Intelligibility classification**

- **R** ~ *Bernoulli* (ρ)

| **Feature set** | **Logistic regression** | **Reliability aware** | **Avg. Reliability** |
|---|---|---|---|
| voice quality | 58.2 | **59.8** | **0.43** |
| prosody | 67.1 | 66.7 | 0.73 |
| pronunciation | 55.1 | **56.2** | **0.16** |
| feature fusion | 68.0 | 67.8 | 0.78 |

Reliability aware model improves classification when feature set is less reliable

# Summary

## Pros

- discriminative modeling allows for reliable parameter estimation
- learns regions in feature space where annotations are more reliable

## Cons

- linear class boundary for reliability in feature space may not be ideal
- model is unable to combine reliable information from different feature subsets

# Questions?