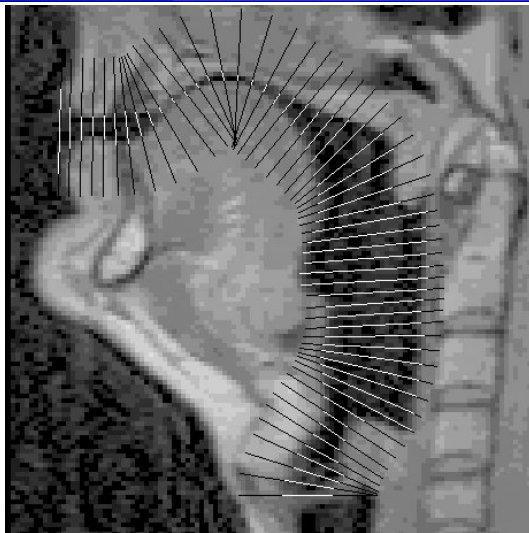


# A Summer at Johns Hopkins

June–August 2008  
Baltimore, MD



**Stephen Shum**

EECS Undergraduate, 2009

University of California, Berkeley

ICSI Lunch Talk, October 14, 2008

# The Journey



- Introduction
- Other Teams
  - Multilingual Spoken Term Detection
  - Robust Speaker Recognition
- My Team
  - Vocal Aging Explained by Vocal Tract Modeling
- Wrap Up

# The Journey



- Introduction
- Other Teams
  - Multilingual Spoken Term Detection
  - Robust Speaker Recognition
- My Team
  - Vocal Aging Explained by Vocal Tract Modeling
- Wrap Up

# Introduction – About the Workshop



- Internationally noted event where mixed teams of leading professionals and students collaborate to advance the state-of-the-art in speech and language technologies.
- Led by:
  - Professor Fred Jelinek
  - Associate Professor Sanjeev Khudanpur

# Introduction – About Johns Hopkins



- Center for Language and Speech Processing (CLSP)
- Johns Hopkins University, Baltimore, Maryland





# Introduction



- 3 Teams:
  - Multilingual Spoken Term Detection
    - Finding and Testing new pronunciations from the web
  - Robust Speaker Recognition
    - Improving state-of-the-art Factor Analysis methodology
  - Vocal Aging Explained
    - How does a person's voice change with age?

# The Journey



- Introduction
- Other Teams
  - Multilingual Spoken Term Detection
    - Led by Professor Richard Sproat (UIUC)
    - Thanks to Erica Cooper (MIT) and Kristy Hollingshead (OGI) for help with the slides.
  - Robust Speaker Recognition
- My Team
  - Vocal Aging Explained by Vocal Tract Modeling
- Wrap Up

# Multilingual Spoken Term Detection



- Extracting and evaluating “found” pronunciations
  - Pronunciations using the International Phonetic Alphabet  
Lorraine Albright /'ɔl brait/
  - Ad-hoc Pronunciations  
bruschetta (pronounced broo-SKET-uh)
  - Worked on ways to filter out “noise” from the Internet.
    - “...Stephen, pronounced clinically insane...”



# Multilingual Spoken Term Detection



- Enhancing ASR with new sources of pronunciation
  - Test whether these new pronunciations help us find more spoken instances of these words in news broadcasts.
  - A good evaluation metric for pronunciations:
    - If some pronunciation doesn't get you better recall, then it's probably not a good pronunciation to use.

# The Journey



- Introduction
- Other Teams
  - Multilingual Spoken Term Detection
  - Robust Speaker Recognition
    - Led by Professor Lukas Burget (Brno)
    - Thanks to Jason Pelecanos (IBM) for the slides
- My Team
  - Vocal Aging Explained by Vocal Tract Modeling
- Wrap Up

# The Joint Factor Analysis Model



- Probabilistic model proposed by Patrick Kenny
- Speaker model represented by mean supervector

$$\mathbf{M} = \mathbf{m} + \mathbf{V}\mathbf{y} + \mathbf{D}\mathbf{z} + \mathbf{U}\mathbf{x}$$

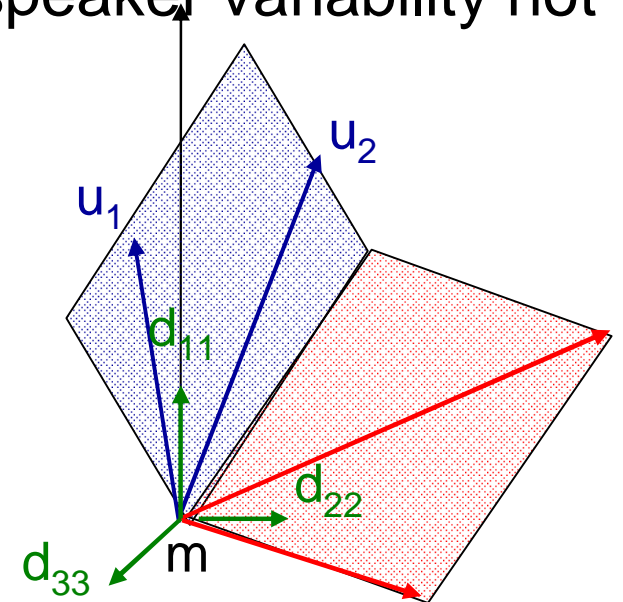
$\mathbf{U}$  – subspace with high intersession/channel variability (eigenchannels)

$\mathbf{V}$  – subspace with high speaker variability (eigenvoices)

$\mathbf{D}$  – diagonal matrix describing remaining speaker variability not covered by  $\mathbf{V}$

Gaussian priors assumed for speaker factors  $\mathbf{y}$ ,  $\mathbf{z}$ , and channel factors  $\mathbf{x}$

3D space of model parameters  
(e.g. 3 component GMM; 1D features)



# Robust Speaker Recognition



## ■ Factor Analysis Conditioning

### ■ Problem

- A single FA model is sub-optimal across different conditions
  - (e.g. duration, phonetic content, recording scenario)

### ■ Two Approaches

- Build FA models specific to each condition,
  - **Robustly combine** multiple models
- Extend the FA model to **explicitly model** these conditions
  - Just another source of variability

# Robust Speaker Recognition



## ■ Factor Analysis Conditioning

### ■ Results and Outcomes

- Showed robustness using within-session variability modeling

$$M = m + Vy + Dz + Ux \rightarrow M = m + Vy + Dz + U_1x + U_w w$$

$U$  – subspace with high intersession/channel variability (eigenchannels)

$U_1$  – intersession part

$U_w$  – within-session part

$V$  – subspace with high speaker variability (eigenvoices)

$D$  – diagonal matrix describing remaining speaker variability not covered by  $V$

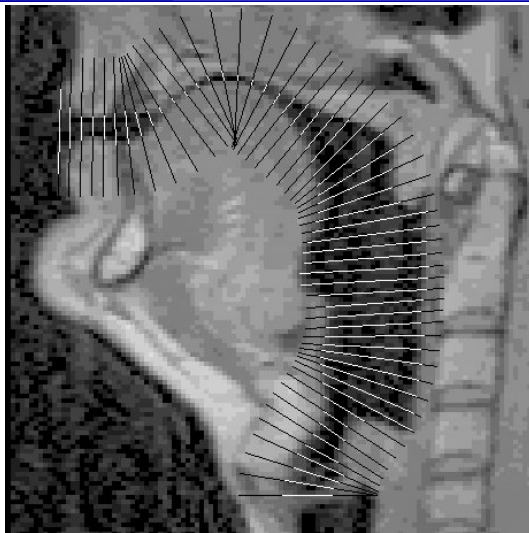
# The Journey



- Introduction
- Other Teams
  - Multilingual Spoken Term Detection
  - Robust Speaker Recognition
- My Team
  - Vocal Aging Explained by Vocal Tract Modeling
    - Led by Professor Elmar Noeth (Univ. Erlangen)
- Wrap Up



# Vocal Aging Explained by Vocal Tract Modeling



**Peter Beyerlein, Andrew Cassidy, Varada Kholhatkar, Eva Lasarcyk, Elmar Nöth, Blaise Potard, Stephen Shum, Young Chol Song, Werner Spiegl, Georg Stemmer, Puyang Xu**

**Special thanks to**

**Andreas Andreou and Sanjeev Khudanpur**

Baltimore, August 13th 2008

# Roadmap



## ■ Problem, Data

- Human performance
- Source
  - A model of the glottis excitation
- Filter
  - Articulatory inversion
- Parameters
  - Duration and speaking rate
  - Other acoustic and prosodic parameters
- Age prediction
- Speech recognition in the context of aging
  
- Outlook & Summary

# What we know



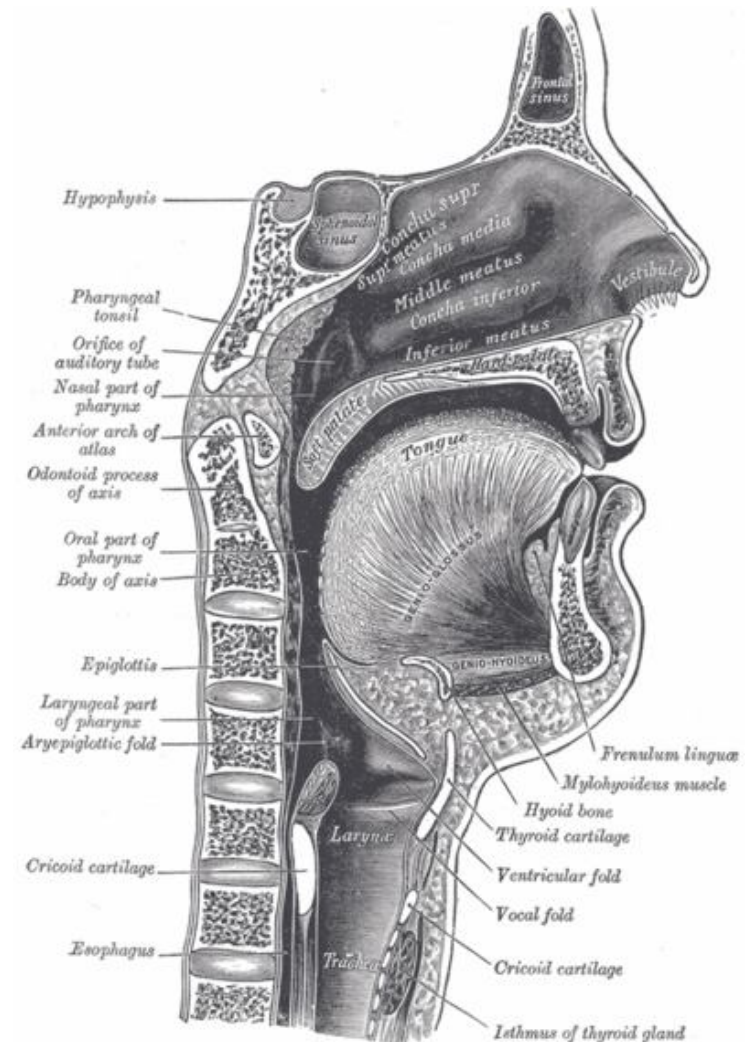
- A person's voice changes due to...
  - Aging
  - Emotional conditions
  - Pathological conditions



# What we see



- Physiological changes
- Acoustic changes
- Somewhat inconclusive findings in the literature



# What we set out to do



- Analyze influence of age on speech:
  - Human perception
  - Objective measurements
  
- Applications
  - Age adaptation for improved ASR
  - Classify the age of a person from his/her speech

# Longitudinal Data



Two British English (BE) speakers: data from several years over roughly a 50 year period



- **Queen Elizabeth II**

- b. 1926, accent = RP, Christmas broadcasts, 30 recordings (1952-2002), 5-10 min each → 2.5 h



- **Alistair Cooke**

- b. 1908, accent = RP with N. American influences, 'Letter from America', 30 recordings (1947 – 2003), ~25 min each → 10.6 h



# Cross-Sectional Data

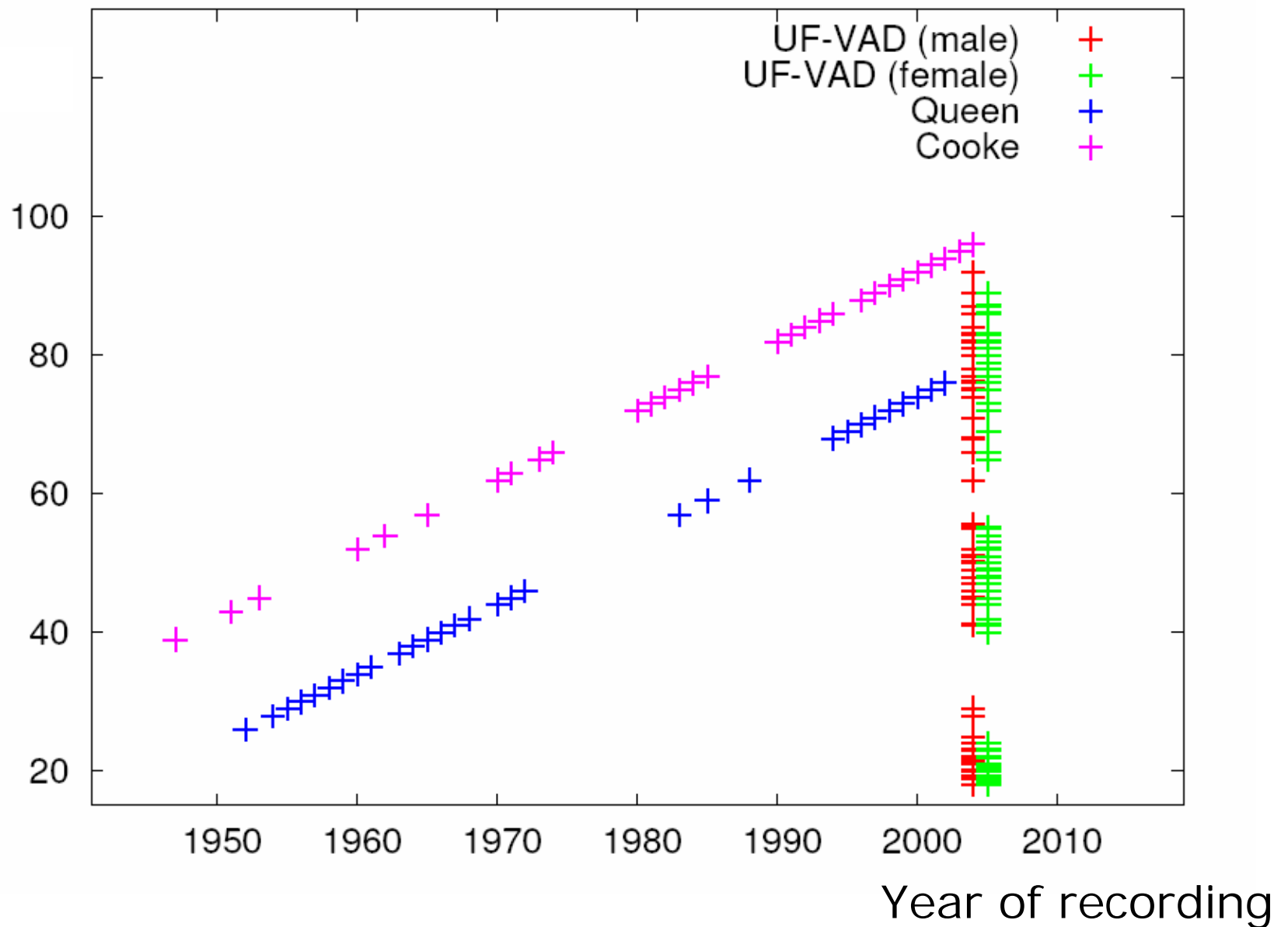


Source:	University of Florida Vocal Aging Database
Language:	American English (AE)
Description:	-25 male, 25 female -3 chronological groups (old, middle-aged, young)
Notes:	-Same text -Read speech -Recorded on <u>same</u> microphone -5 hrs total

# Longitudinal vs. Cross-sectional Data



Age



# Roadmap



- Data
- **Human performance**
- Source
  - A model of the glottis excitation
- Filter
  - Articulatory inversion
- Parameters
  - Duration and speaking rate
  - Other acoustic and prosodic parameters
- Age prediction
- Speech recognition in the context of aging
- Outlook & Summary

# Vocal Aging Explained ...

## ...by Humans



User Registration, Login - Mozilla Firefox

http://www.dsp.jhu.edu/vaevtmapp/test1/

Erste Schritte Aktuelle Nachrichten

Google Suche

Vocal Aging Lab - CLSP Wiki

### Vocal Aging Explained

#### by Vocal Tract Modeling

Please enter your information in the form below\*\*:

Name:

Email Address:

Gender: ☐ Male ☐ Female

Age:

First Language:  If Other, please specify:

English Level (5=highest/native):

Briefly explain, if desired:

(Optional) Affiliation to Project:

If Other, please specify:

\*\*Privacy Note: The information that you submit here and in the subsequent pages will only be seen by researchers of the 2008 Summer Workshop at the Center for Language and Speech Processing

# Vocal Aging Explained ...

## ...by Humans



- 30 recordings of the Queen of England
  - ~10 seconds in length
  - Hand-selected → absence of history-sensitive words (i.e. wars, political situations, natural phenomena, etc.)

# The Website



Question 1 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.cslp.jhu.edu/vaevtmapp/vocalagingtest/questions.php

Google

Latest Headlines Berkeley Weather Barack Obama's Spee...

## Vocal Aging Explained

### by Vocal Tract Modeling

#### Question 1

[LEFT](#) [RIGHT](#)

Please answer, to the best of your ability, the following questions about the two sound samples above.

**Is the recording on the LEFT of an Older or Younger speaker?**

☐ Older ☐ Younger

**By how many years?**

**Please estimate the Age of the speaker in the recording on the LEFT:**  
(values between 20 and 80 are would be most reasonable...)



# The Website



Question 1 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.csp.jhu.edu/vaevtmapp/vocalagingtest/questions.php

Google

Latest Headlines Berkeley Weather Barack Obama's Spee...

## Vocal Aging Explained

### by Vocal Tract Modeling

#### Question 1

[LEFT](#) [RIGHT](#)

Please answer, to the best of your ability, the following questions about the two sound samples above.

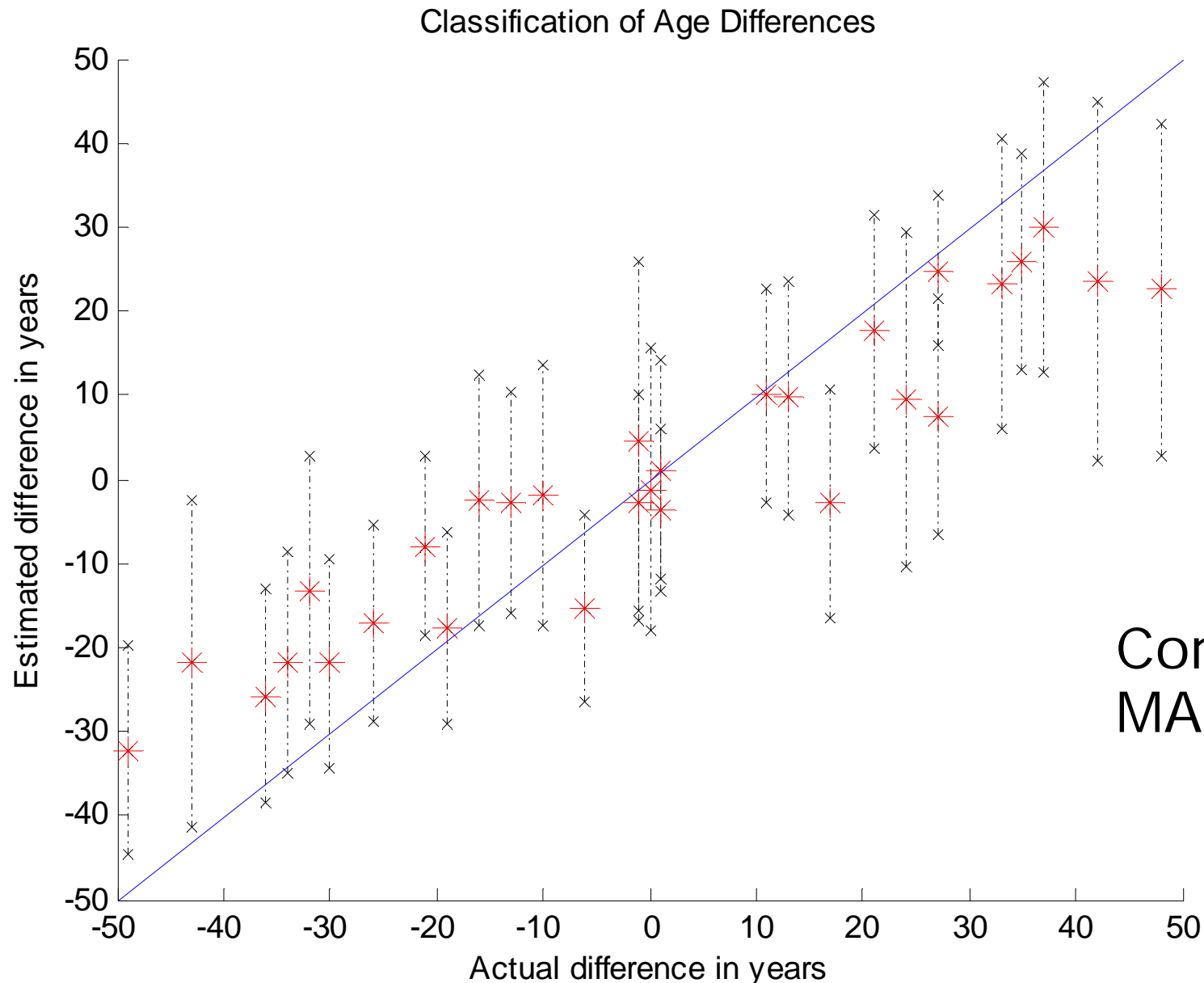
**Is the recording on the LEFT of an Older or Younger speaker?**

☐ Older ☒ Younger

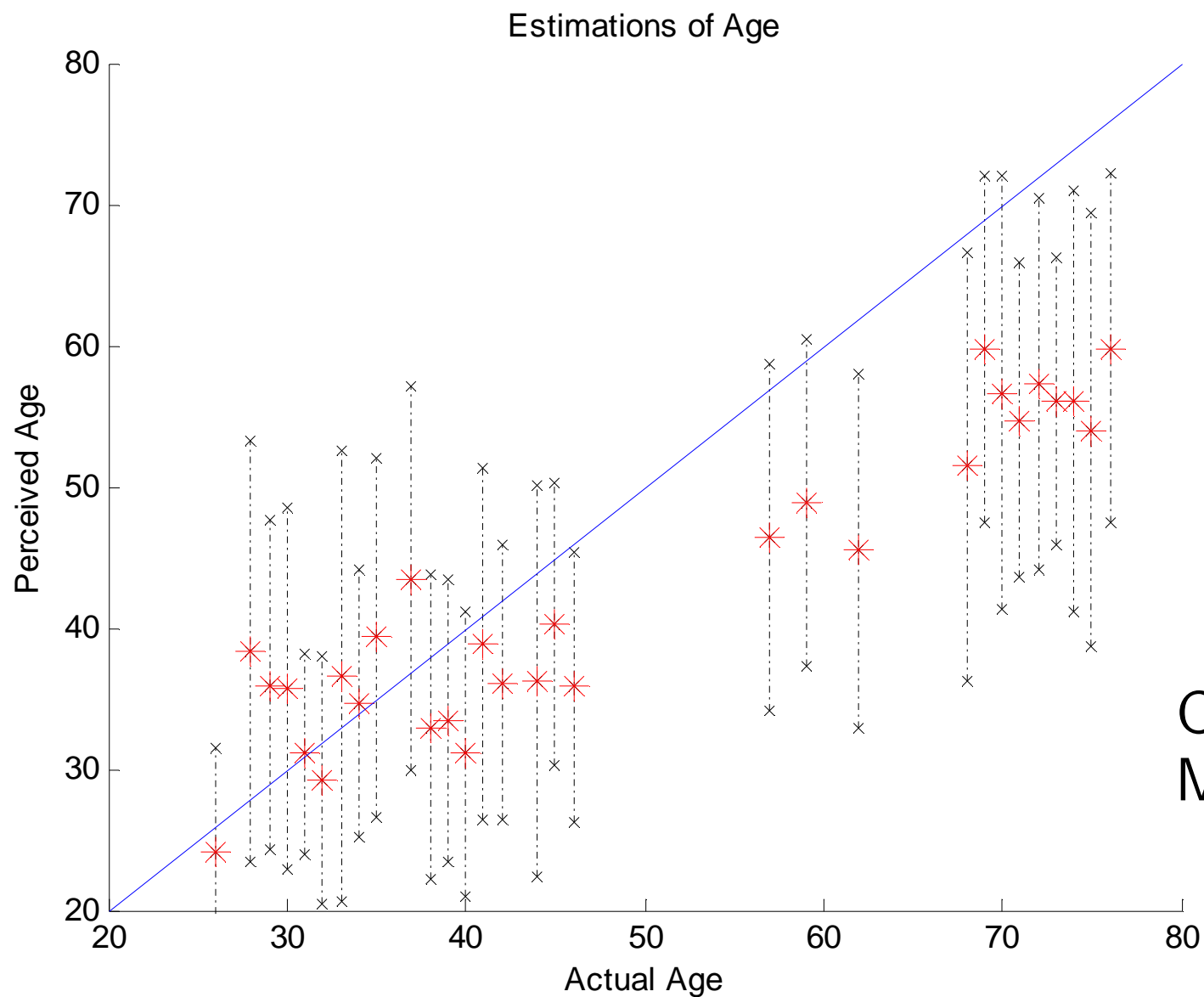
**By how many years?**

**Please estimate the Age of the speaker in the recording on the LEFT:**  
(values between 20 and 80 are would be most reasonable...)

# Current Results (~110 Participants)



# Current Results, cont'd



Corr: 0.93  
MAE: 9.1 Y.

# Potential Bias



## ■ What factors might bias these results?

- Level of proficiency in English
- Age of test taker
- Gender

# Human Age Estimation: Conclusions



- Participants were quite good at classifying
  - age differences (0.94)
  - one's actual age (0.93)
- Seem to underestimate older voices
- No influencing factors such as gender, age, and language proficiency found

# Overview



- Data
- Human performance
- **Source**
  - **A model of the glottis excitation**
- Filter
  - Articulatory inversion
- Parameters
  - Duration and speaking rate
  - Other acoustic and prosodic parameters
- Age prediction
- Speech recognition in the context of aging
- Outlook & Summary

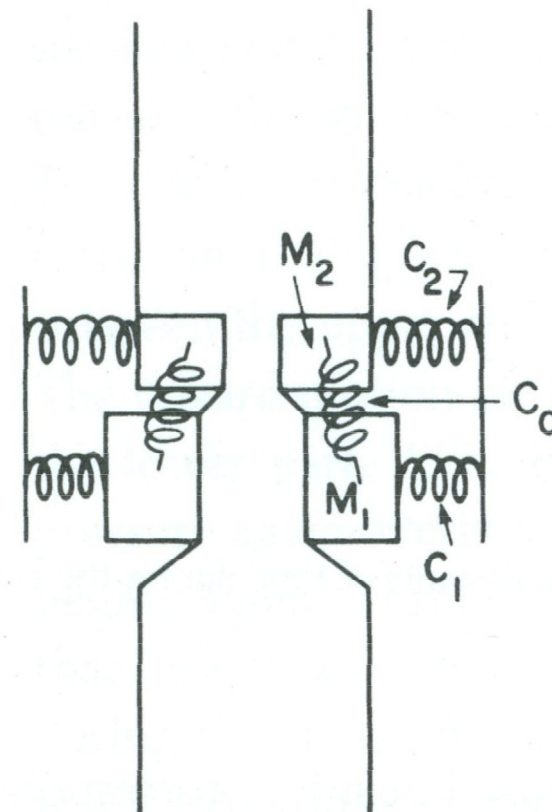
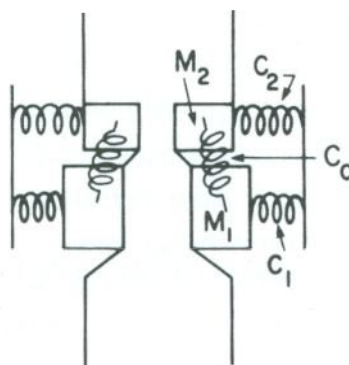
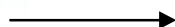
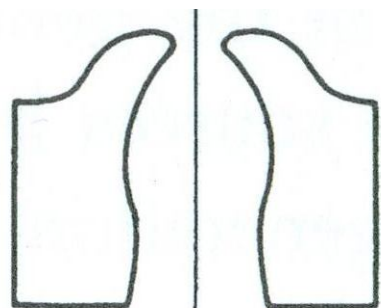


# Glottis Inversion



Goal: Analyze age-related changes in the excitation

- Simple two-mass Vocal Fold Model
  - Kenneth Stevens
- Few parameters => good for data-driven optimization
- Flexible



# Overview



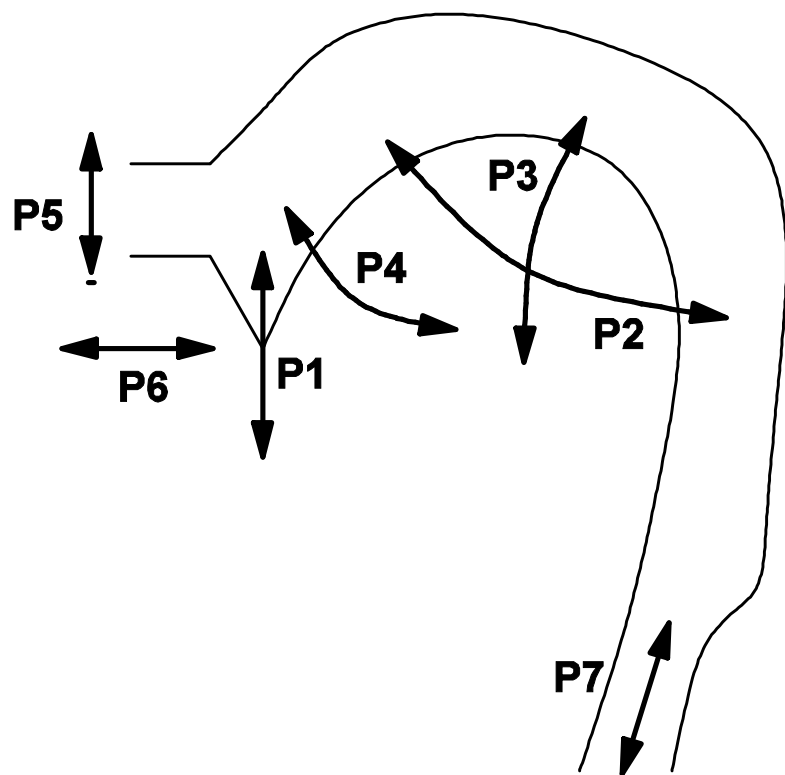
- Data
- Human performance
- Source
  - A model of the glottis excitation
- **Filter**
  - **Articulatory inversion**
- Parameters
  - Duration and speaking rate
  - Other acoustic and prosodic parameters
- Age prediction
- Speech recognition in the context of aging
- Outlook & Summary

# Articulatory Inversion



- Goal
  - (in general) To find the vocal tract shapes from only the speech signal
  - (in this workshop) To study how such articulation changes in the context of aging
- In theory, provides a framework to generate features from the speech signal, channel and source independent.

# Maeda Model



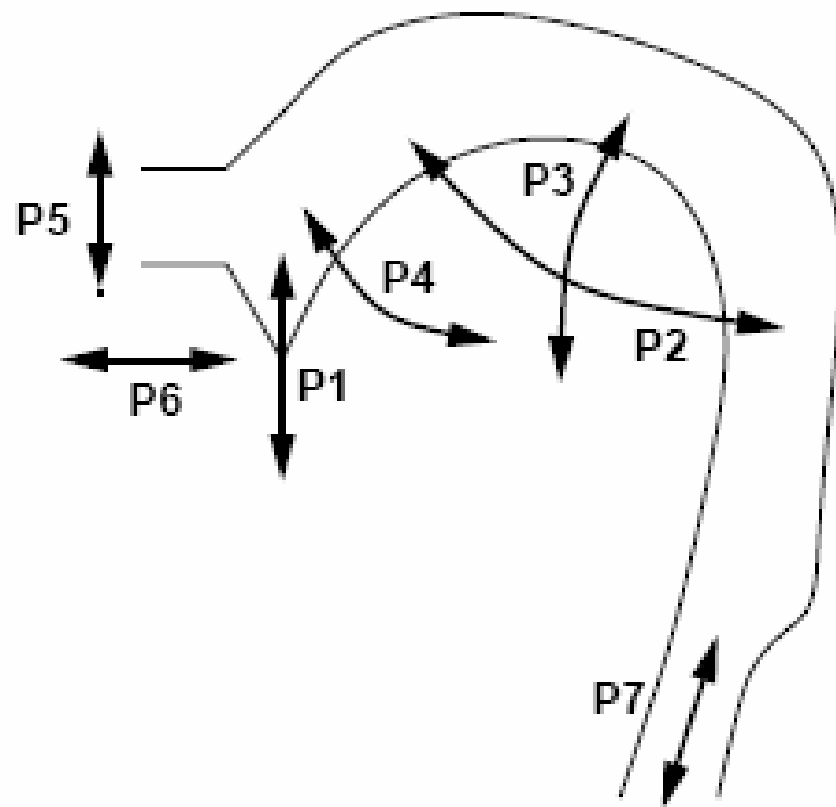
- Derived from X-rays
- 2D sagittal slice model
- 7 deformation modes
- Easy to adapt to different speakers
- Cannot model consonants or nasalized vowels
- Low geometric and acoustic faithfulness

# Maeda Model



## Parameters

- P1 Jaw position
- P2 Tongue dorsum position (front/back)
- P3 Tongue shape (round/flat)
- P4 Tongue tip
- P5 Vertical opening of the lips
- P6 Protrusion of the lips
- P7 Larynx height

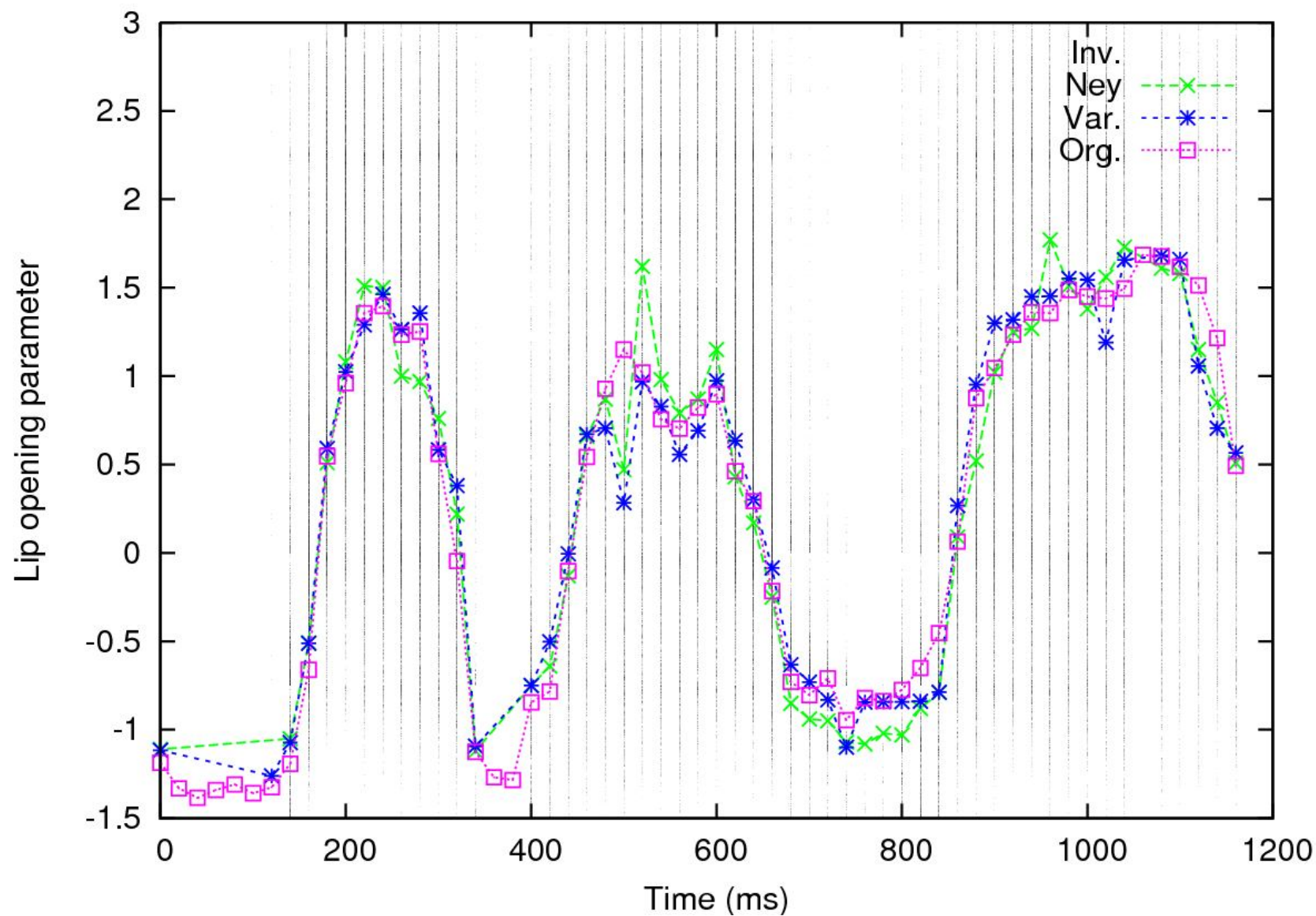


# Inversion process



- Acoustic-to-articulatory Inversion Problem
  - Many-to-one, complex mapping
  
- 3 Steps:
  - Look up possible solutions at each frame
  - Find a trajectory of minimal articulatory cost
  - Improve using variational calculus

# A successful example: P5





## *More Ideas*







# Current Work (Experiments)



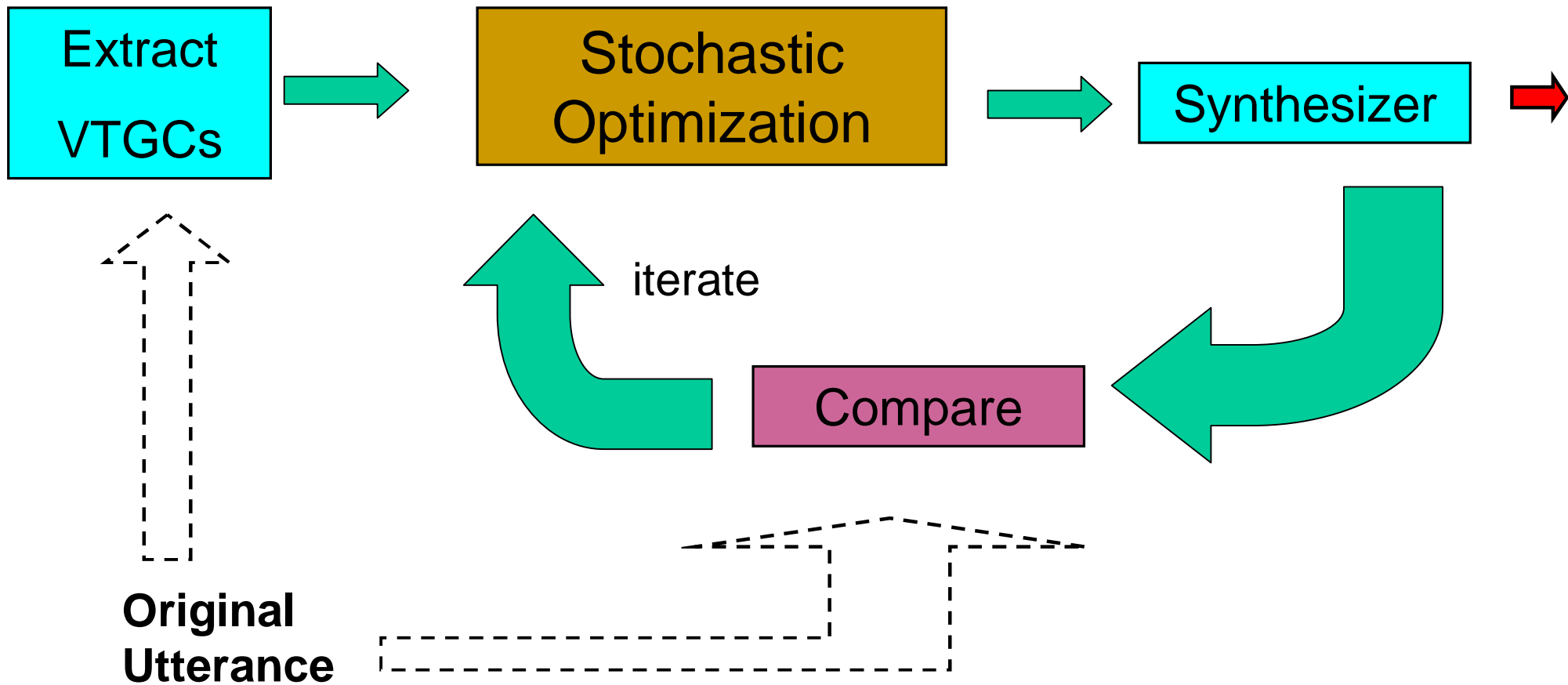
- The 7 articulatory parameters [P1 ... P7] of the Maeda Model inherently resemble features
  - Suggests the possibility of a new feature extraction method
    - Channel independent (use only pitch/formants)
  - We will begin calling these features VTGC's
    - Vocal Tract Geometry Coefficients

# Arising Issues



- Articulatory Parameters [P1 ... P7] (aka VTGCs) do not model consonants or nasalized vowels.
- How well does this synthesizer work?
  - Original: 
  - Resynthesized: 

# The Analysis-by-Synthesis Approach



# Next Steps



- Try to create a codebook of VTGCs for each phoneme.
  - Already exists – MOCHA Database
    - (<http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html>)
- Use only voiced parts of the re-synthesized speech
  - Try doing speech recognition
  - Similar error rates => possible dimensionality reduction
- Open to other suggestions!
  - Improve both the Maeda Model and Re-Synthesizer

# Conclusion



- Found subset of articulatory parameters which are strongly correlated with age
  - Jaw Position, Lip Protrusion, Vertical Lip opening
- The most consistent result was the decrease in amplitude of articulatory movements
- We look to continue work on these articulatory parameters (VTGCs) towards the possibility of a new feature extraction method.

# Roadmap



- Data
- Human performance
- Source
  - A model of the glottis excitation
- Filter
  - Articulatory inversion
- **Parameters**
  - Duration and speaking rate
  - Other acoustic and prosodic parameters
- **Age prediction**
  - Speech recognition in the context of aging
    - Baseline system
    - Age adaptation
    - New approaches
- Outlook & Summary

# Acoustic Investigations

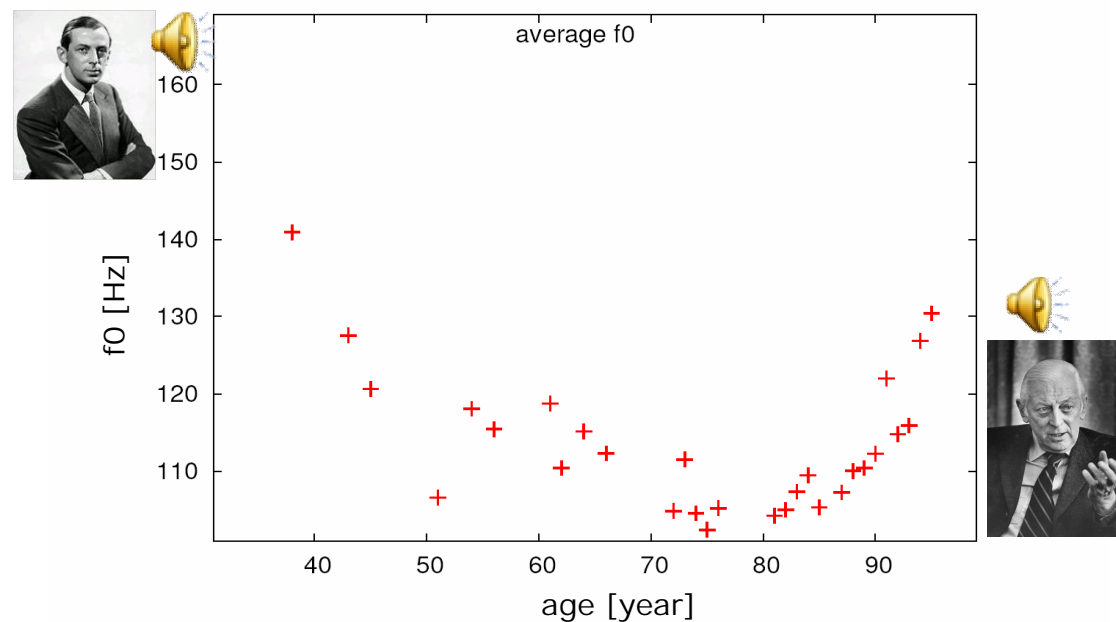
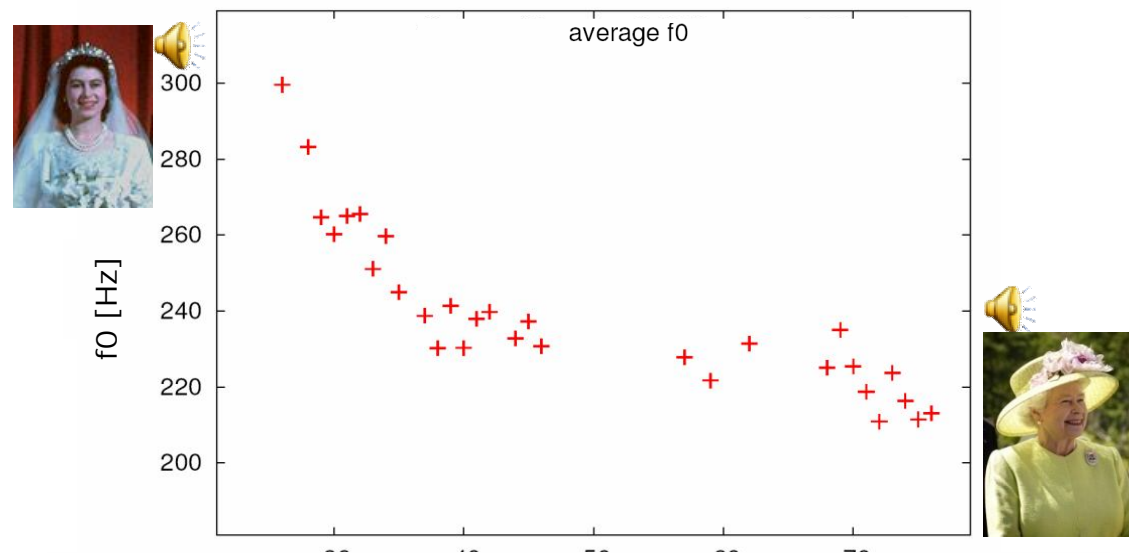


- Speaking Rate - # phones per second
  - Florida male recordings have a -0.74 correlation with age.
  - Not consistent in Queen (professionalism?)
  
- Pause Percentage, Plosive-vowel durations
  - No consistent findings throughout; more work to be done.

# Basic Features – F0



- Frame based
  - Averaged per year
  - Obvious trend on longitudinal data
- Promising feature for speaker dependent experiments





# Age Prediction: Human vs. System



## ■ Humans (average over 40 people)

		Actual Age		
		~35	35~59	59~
Predicted Age	~35	48	4	0
	35~59	2	46	14
	59~	0	0	36

**Accuracy**  
**87%**

**9 years off**  
on average

## ■ Our System

		Actual Age		
		~ 35	35 ~ 59	59 ~
Predicted Age	~ 35	29	1	0
	35 ~ 59	20	43	8
	59 ~	1	6	42

**Accuracy**  
**76%**

**11 years off**  
on average

# Roadmap



- Data
- Human performance
- Source
  - A model of the glottis excitation
- Filter
  - Articulatory inversion
- Parameters
  - Duration and speaking rate
  - Other acoustic and prosodic parameters
- Age prediction

## ■ Speech recognition in the context of aging

- Outlook & Summary

### Vocal Aging Explained by Vocal Tract Modeling



Peter Beyerlein, Andrew Cassidy, Varada Kholhatkar, Eva Lasarcyk, Elmar Nöth, Blaise Potard, Stephen Shum, Young Chol Song, Werner Spiegl, Georg Stemmer, Puyang Xu

Special thanks to  
Andreas Andreou and Sanjeev Khudanpur

Baltimore, August 13th 2008

# Buildup of Broadcast News Speech Recognizer



## ■ Why a Broadcast News system ?

F-Conditions, Robustness, Flexibility help to provide Services on various tasks in the Vocal Aging project

- new feature extraction
- new training
- new decoding
- Based on JHU htk/h4 recipes

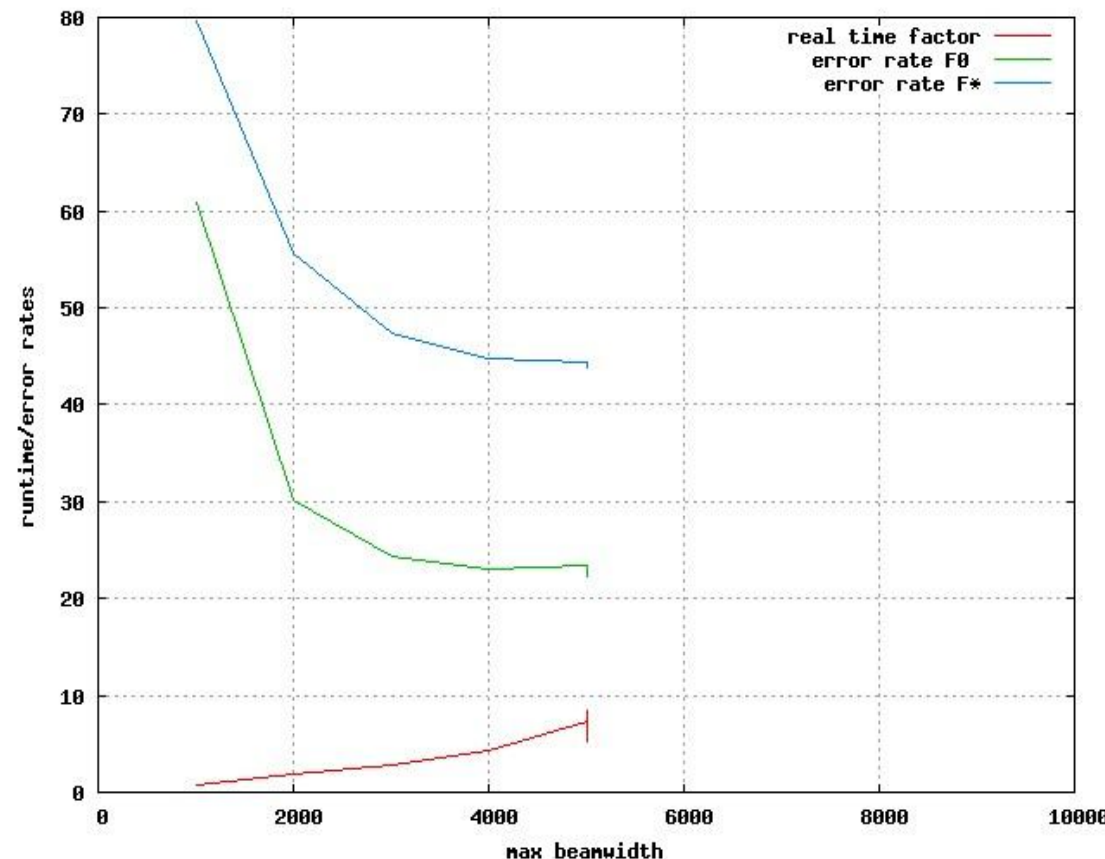
## ■ One pass 3-gram

Xword-3-phone (7xRT)

## ■ H4Dev96

1999 system – 35.8% (NIST Scoring)

2008 system – 34.6% (NIST Scoring)



# Overview



- **Does Age Adaptation help ASR performance ?**
- What happens when Speaker Age mismatches Model Age ?
- Can we predict the future of an Adaptation Transform ?  
And why should we do this ?

## Adaptation to dev data, decoding on eval data



Word Error Rate in %	No Adapt				
Age 26-31	62.0				
Age 31-39	52.6				
Age 40-46	51.1				
Age 57-68	46.0				
Age 68-72	48.9				
Age 72-76	45.6				
<b>Total</b>	<b>51.3</b>				

## Adaptation to dev data, decoding on eval data



Speaker  
Adaptation

Word Error Rate in %	No Adapt		Supervised Adapt to all Dev		
Age 26-31	62.0		40.1 <sub>(-35%)</sub>		
Age 31-39	52.6		37.3 <sub>(-29%)</sub>		
Age 40-46	51.1		32.9 <sub>(-36%)</sub>		
Age 57-68	46.0		30.3 <sub>(-34%)</sub>		
Age 68-72	48.9		31.9 <sub>(-35%)</sub>		
Age 72-76	45.6		25.8 <sub>(-43%)</sub>		
<b>Total</b>	<b>51.3</b>		<b>33.3</b> <b>(-35%)</b>		

## Adaptation to dev data, decoding on eval data



Speaker  
Adaptation

Age Adaptation

Word Error Rate in %	No Adapt		Supervised Adapt to all Dev	Supervised Adapt to Age	
Age 26-31	62.0		40.1 <sub>(-35%)</sub>	37.3 <sub>(-40%)</sub>	
Age 31-39	52.6		37.3 <sub>(-29%)</sub>	36.7 <sub>(-30%)</sub>	
Age 40-46	51.1		32.9 <sub>(-36%)</sub>	28.9 <sub>(-43%)</sub>	
Age 57-68	46.0		30.3 <sub>(-34%)</sub>	30.3 <sub>(-34%)</sub>	
Age 68-72	48.9		31.9 <sub>(-35%)</sub>	29.1 <sub>(-40%)</sub>	
Age 72-76	45.6		25.8 <sub>(-43%)</sub>	24.3 <sub>(-47%)</sub>	
<b>Total</b>	<b>51.3</b>		<b>33.3</b> <b>(-35%)</b>	<b>31.4</b> <b>(-39%)</b>	

# Adaptation to dev data, decoding on eval data



		Random Subset Adaptation	Speaker Adaptation	Age Adaptation	
Word Error Rate in %	No Adapt	Supervised Adapt to Random Chunks	Supervised Adapt to all Dev	Supervised Adapt to Age	
Age 26-31	62.0	44.2 <sub>(-29%)</sub>	40.1 <sub>(-35%)</sub>	37.3 <sub>(-40%)</sub>	
Age 31-39	52.6	41.6 <sub>(-21%)</sub>	37.3 <sub>(-29%)</sub>	36.7 <sub>(-30%)</sub>	
Age 40-46	51.1	33.6 <sub>(-34%)</sub>	32.9 <sub>(-36%)</sub>	28.9 <sub>(-43%)</sub>	
Age 57-68	46.0	30.7 <sub>(-33%)</sub>	30.3 <sub>(-34%)</sub>	30.3 <sub>(-34%)</sub>	
Age 68-72	48.9	32.4 <sub>(-34%)</sub>	31.9 <sub>(-35%)</sub>	29.1 <sub>(-40%)</sub>	
Age 72-76	45.6	30.4 <sub>(-33%)</sub>	25.8 <sub>(-43%)</sub>	24.3 <sub>(-47%)</sub>	
<b>Total</b>	<b>51.3</b>	<b>35.9</b> <b>(-30%)</b>	<b>33.3</b> <b>(-35%)</b>	<b>31.4</b> <b>(-39%)</b>	

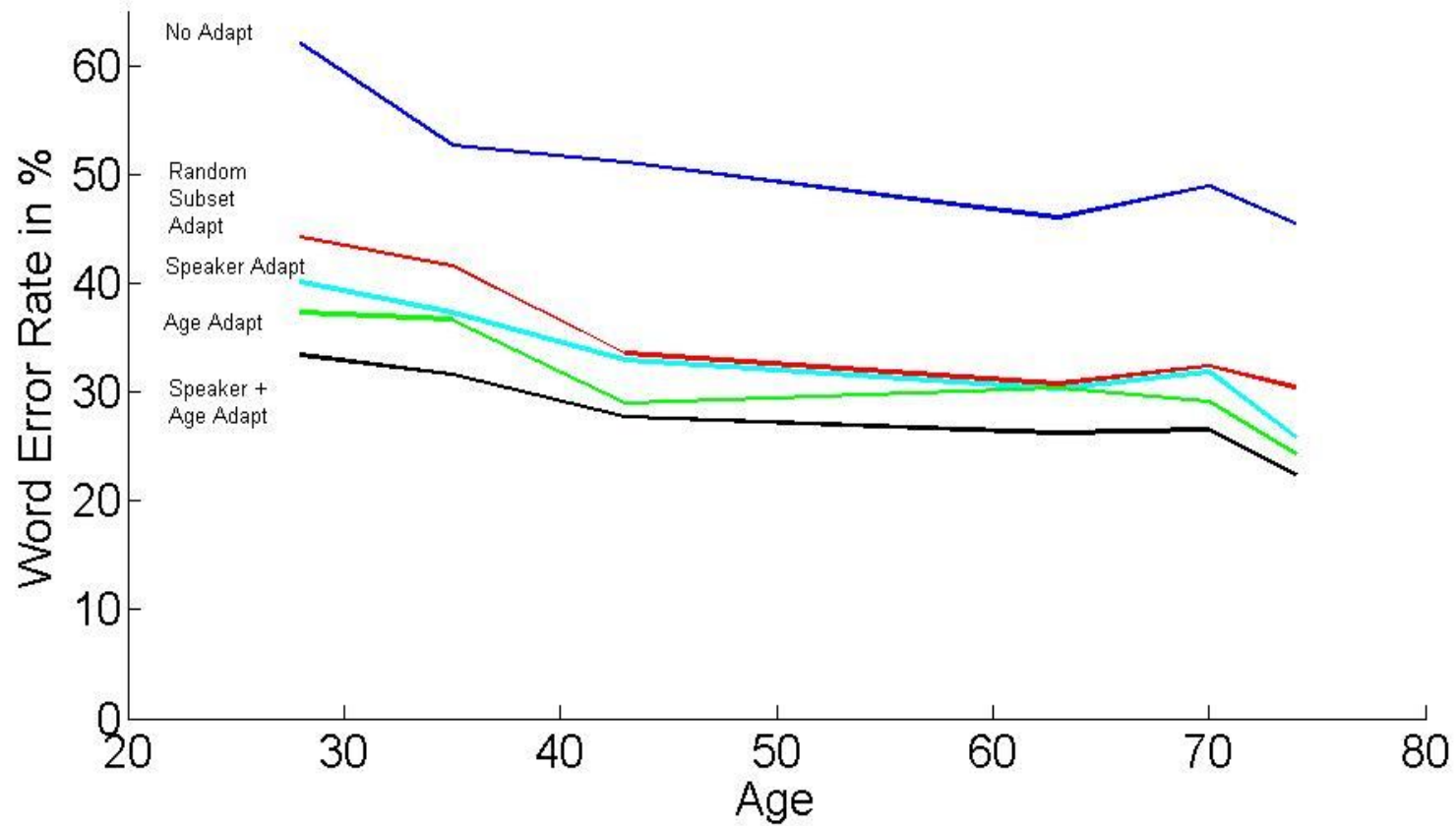


## Adaptation to dev data, decoding on eval data



		Random Subset Adaptation	Speaker Adaptation	Age Adaptation	Speaker + Age Adaptation
Word Error Rate in %	No Adapt	Supervised Adapt to Random Chunks	Supervised Adapt to all Dev	Supervised Adapt to Age	Supervised Adapt to All dev + Supervised Adapt to Age
Age 26-31	62.0	44.2 <sub>(-29%)</sub>	40.1 <sub>(-35%)</sub>	37.3 <sub>(-40%)</sub>	33.4 <sub>(-46%)</sub>
Age 31-39	52.6	41.6 <sub>(-21%)</sub>	37.3 <sub>(-29%)</sub>	36.7 <sub>(-30%)</sub>	31.6 <sub>(-40%)</sub>
Age 40-46	51.1	33.6 <sub>(-34%)</sub>	32.9 <sub>(-36%)</sub>	28.9 <sub>(-43%)</sub>	27.7 <sub>(-45%)</sub>
Age 57-68	46.0	30.7 <sub>(-33%)</sub>	30.3 <sub>(-34%)</sub>	30.3 <sub>(-34%)</sub>	26.2 <sub>(-43%)</sub>
Age 68-72	48.9	32.4 <sub>(-34%)</sub>	31.9 <sub>(-35%)</sub>	29.1 <sub>(-40%)</sub>	26.5 <sub>(-46%)</sub>
Age 72-76	45.6	30.4 <sub>(-33%)</sub>	25.8 <sub>(-43%)</sub>	24.3 <sub>(-47%)</sub>	22.5 <sub>(-51%)</sub>
<b>Total</b>	<b>51.3</b>	<b>35.9</b> <b>(-30%)</b>	<b>33.3</b> <b>(-35%)</b>	<b>31.4</b> <b>(-39%)</b>	<b>28.2</b> <b>(-45%)</b>

## Adaptation to dev data, decoding on eval data



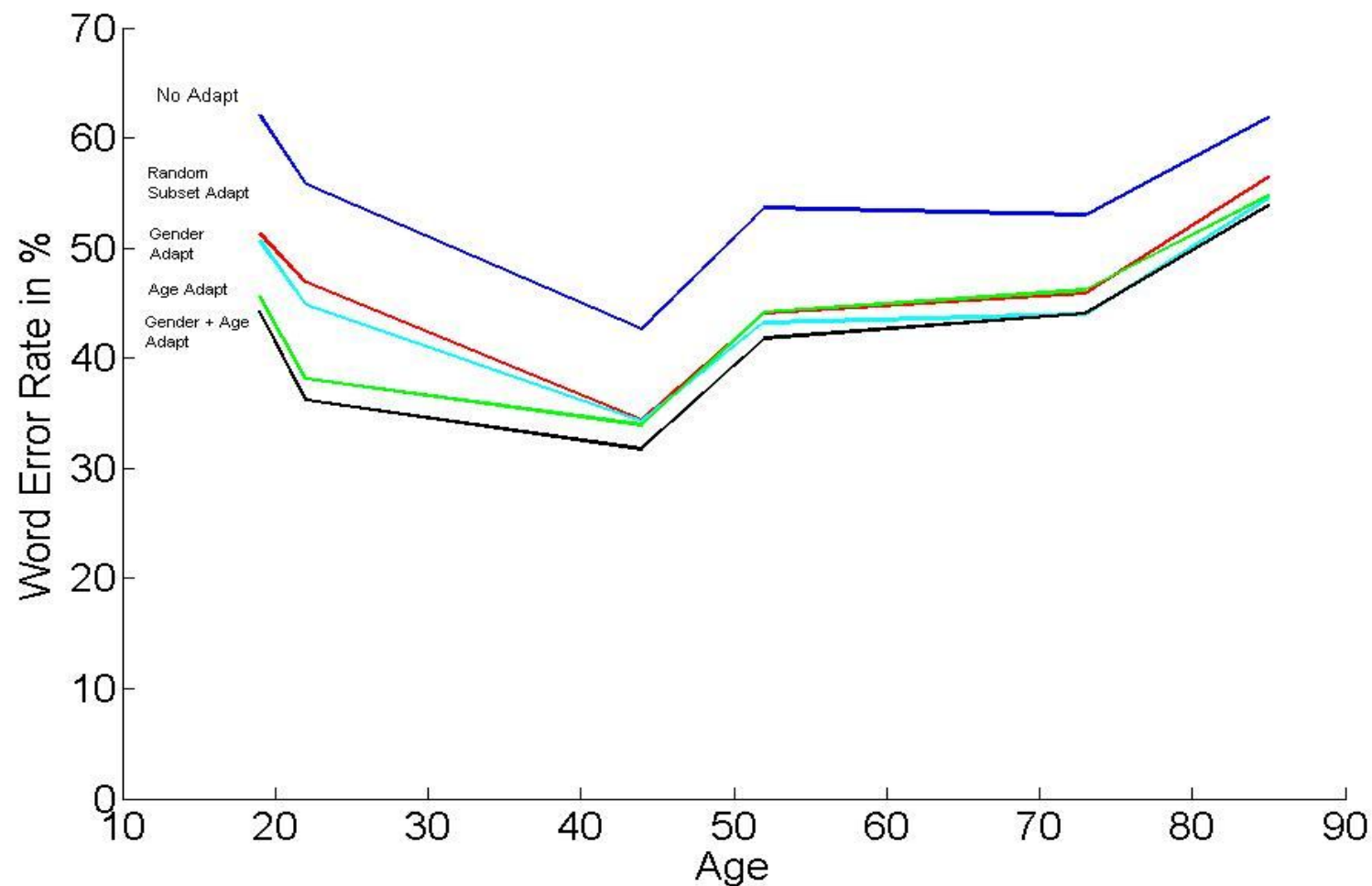
# UF\_VAD (Florida) Corpus



- A speaker independent experiment

Word Error Rate in %	No Adapt	Supervised Adapt to Random Chunks	Supervised Adapt to all female Dev	Supervised Adapt to Age	Supervised Adapt to all female and age
Total	54.8	46.5 (-15%)	45.3 (-17%)	43.8 (-20%)	42.0 (-23%)

# Adapting to female dev data, decoding on female eval data

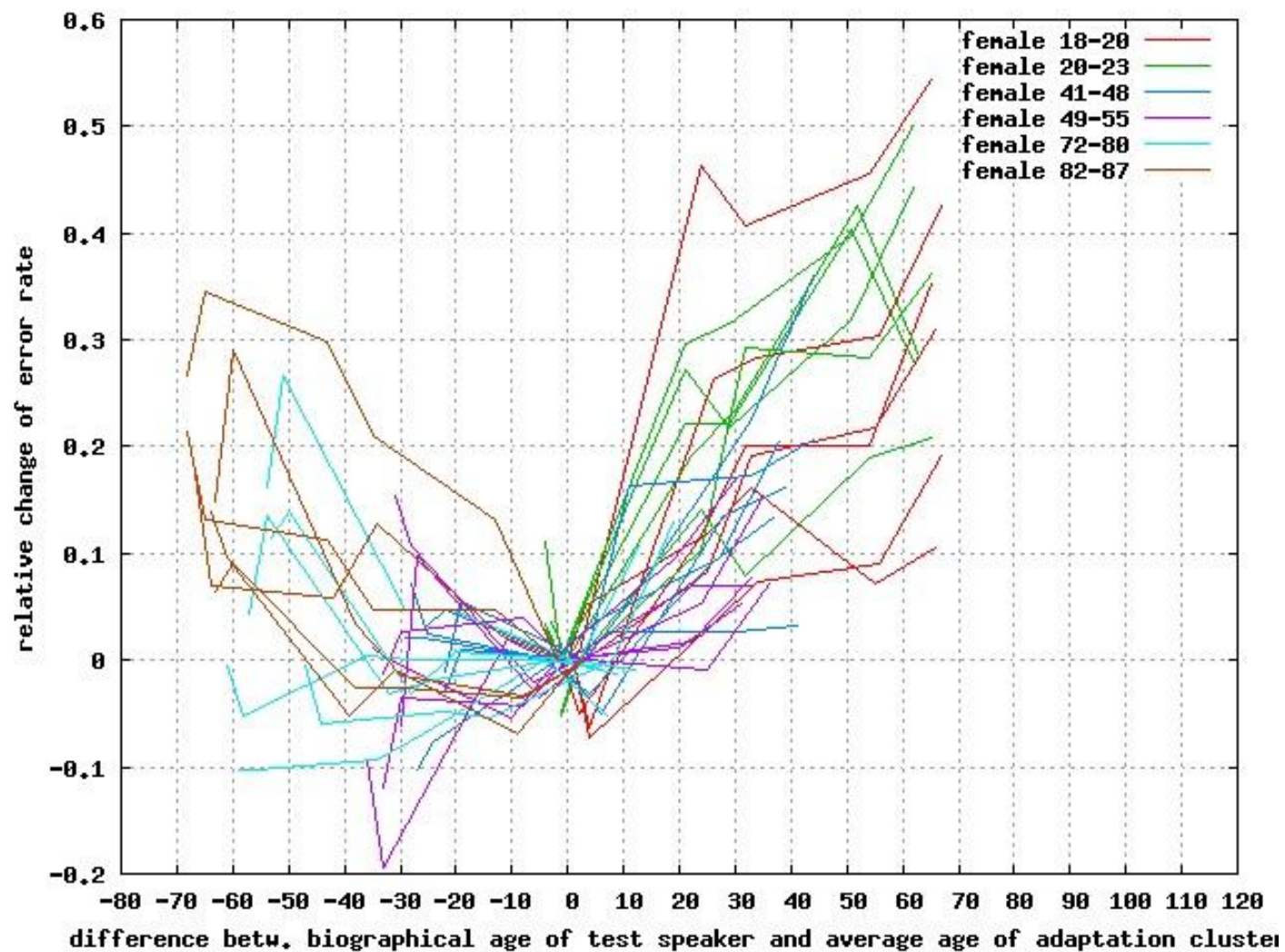


# Overview



- Does Age Adaptation help ASR performance ?
- **What happens when Speaker Age mismatches Model Age ?**
- Can we predict the future of an Adaptation Transform ?  
And why should we do this ?

# A Butterfly Plot



# Overview



- Does Age Adaptation help ASR performance ?
- What happens when Speaker Age mismatches Model Age ?
- **Can we predict the future of an Adaptation Transform ?**  
**Why should we do this ?**
  - Time Permitting...

# Predicting the future of an MLLR



- **Defense**

- Predict optimal ASR models for a target (5 years ahead)  
i.e. be prepared for future changes of the targets voice

- **Medical**

- Predict future ASR models of a patient based on (an assumed) healthy development
- Act if true models start to deviate from expected

- **Society**

- Build ASR systems for elderly people that age together with their users



# Interpolated Age Adaptation on Queen



WER in %	Adapt to 31-39		Adapt to Random Chunk		
Age 31-39	36.65		41.55		

WER in %	Adapt to 68-72		Adapt to Random Chunk		
Age 68-72	29.13		32.38		

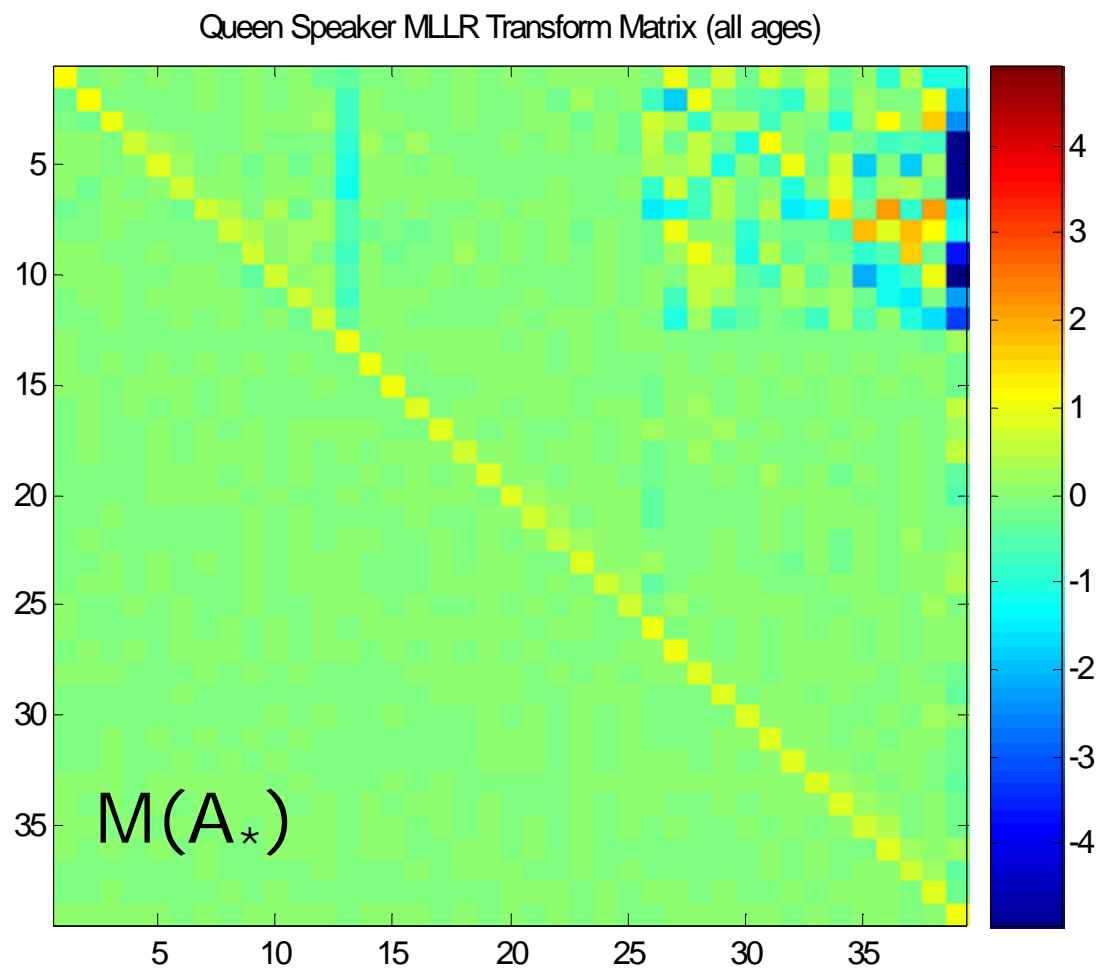
# Interpolated Age Adaptation on Queen



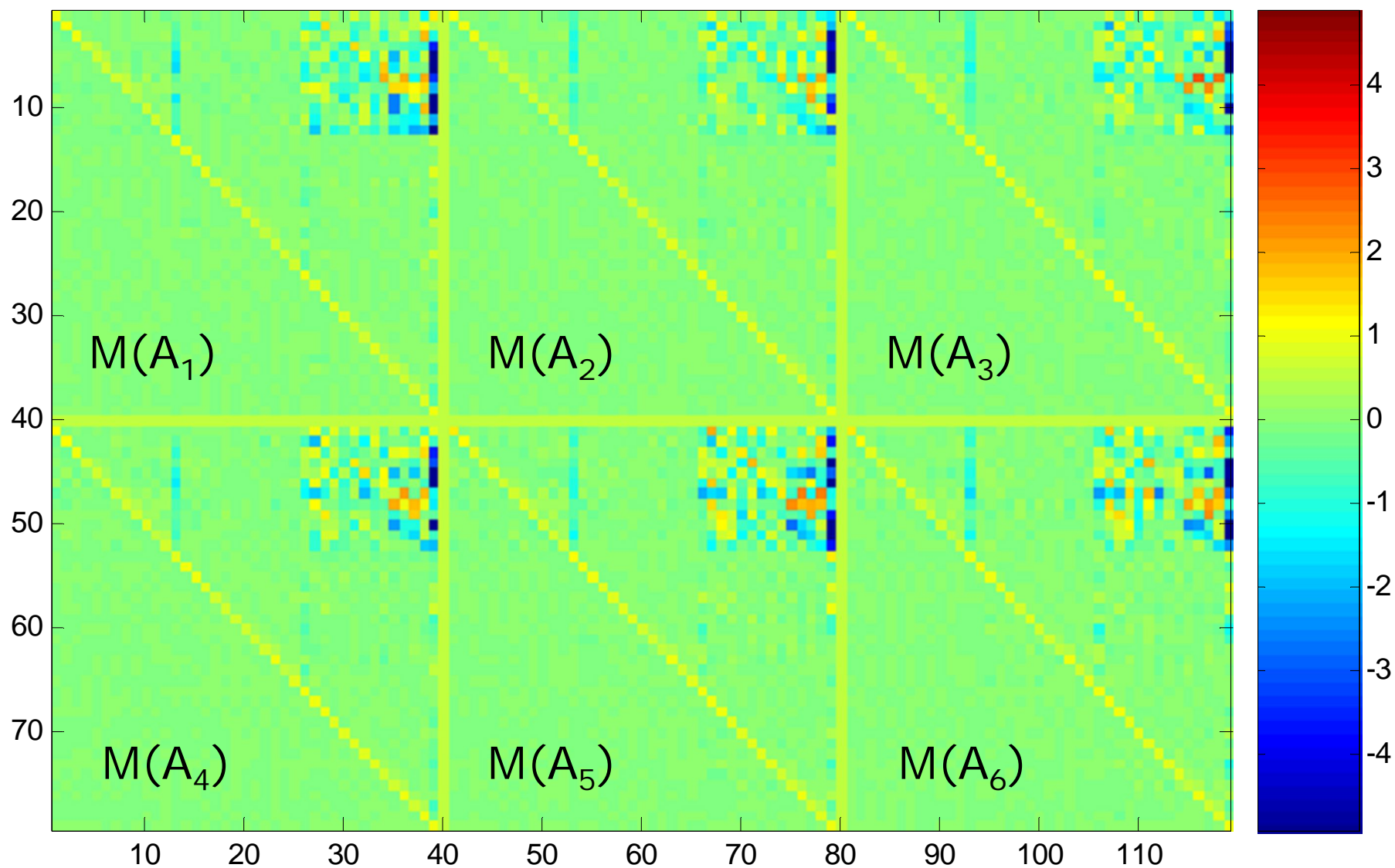
WER in %	Adapt to 31-39	<b>Adapt to 26-31 &amp; 40-46</b>	Adapt to Random Chunk	Adapt to 26-31	Adapt to 40-46
Age 31-39	36.65	<b>37.16</b>	41.55	39.02	36.03

WER in %	Adapt to 68-72	<b>Adapt to 57-68 &amp; 72-76</b>	Adapt to Random Chunk	Adapt to 57-68	Adapt to 72-76
Age 68-72	29.13	<b>30.46</b>	32.38	31.45	30.46

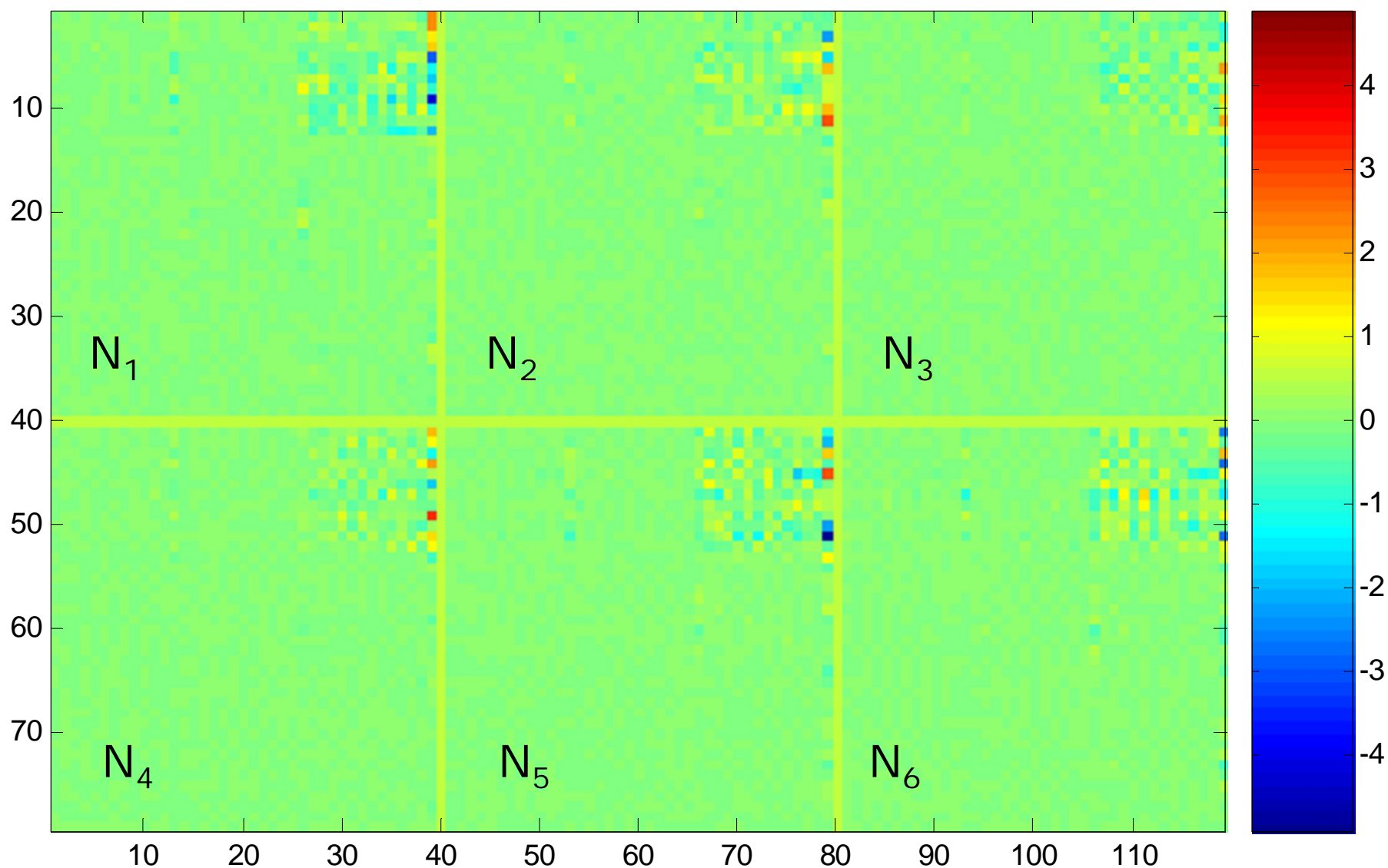
# MLLR Matrix for Queen (all Ages, $A_*$ )



# MLLR Matrices for Age Periods $[A_1, \dots, A_6]$



# Normalized Age MLLR $[N_i = M(A_i) - M(A_*) , i = 1:6]$



## Further Work



- Really understand what is going on in these plots
- Develop methods that can predict an MLLR-future given an MLLR-past

# Roadmap



- Data
- Human performance
- Source
  - A model of the glottis excitation
- Filter
  - Articulatory inversion
- Parameters
  - Duration and speaking rate
  - Other acoustic and prosodic parameters
- Age prediction
- Speech recognition in the context of aging

## ■ Outlook & Summary



# Key Ideas from this Workshop



- Classification of speaker age
  - Approaching human ability
- Excitation signal + Maeda Model parameters
  - Competitive in explaining acoustic changes
- Application to Speech Recognition
  - Established a systematic relationship between speaker age, model age, and resulting WER.
- Pioneered (somewhat) the use of longitudinal data



# Open Questions after the Workshop



- *Can we improve performance by combining MFCCs with model-based parameters in some way?*
- *Can we extend these ideas to other tasks?*
  - *speaker-recognition, emotion-recognition, etc.*
- *Can we find medically relevant deviations from the normal aging process?*
- *Can we utilize MRI and Ultrasound imaging data?*

# Summary



- **Data:** Longitudinal & Cross-sectional
- **Human performance:** Web-based test
- **Source:** Implemented and optimized Steven's model
- **Filter:**
  - Found some articulatory parameters that change with age
- **Acoustic and Prosodic Parameters:**
  - Speaking rate, pauses, F0
- **Age prediction:** mean error  $\leq 8$  years
- **Speech recognition:**
  - Improvement with age adaptation

# The Journey



- Introduction
- Other Teams
  - Multilingual Spoken Term Detection
  - Robust Speaker Recognition
- My Team
  - Vocal Aging Explained by Vocal Tract Modeling
- Wrap Up



# Thanks for listening !

## Questions?