



# Information Visualization, Exploration and Error Analysis in the NIST 2010 Speaker Recognition Evaluation

Xinhui Zhou, **Daniel Garcia-Romero**  
Shihab Shamma and Carol Espy-Wilson

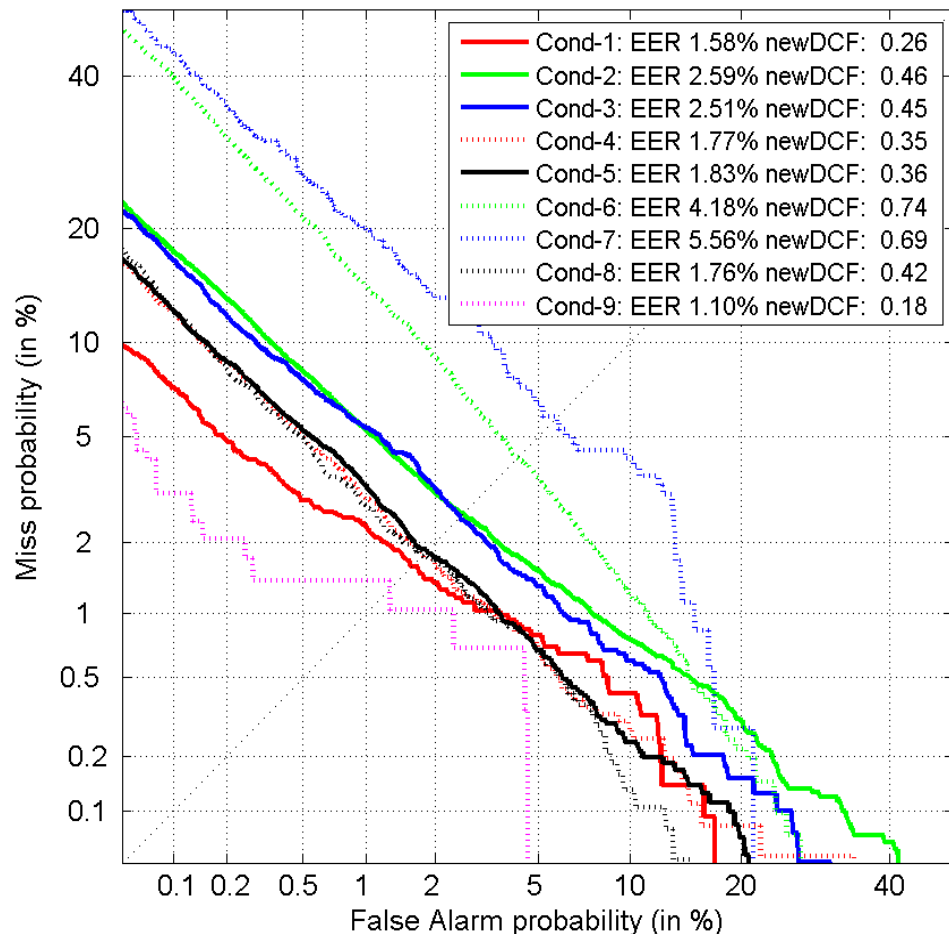
Department of Electrical & Computer Engineering

**University of Maryland, College Park, MD, USA**

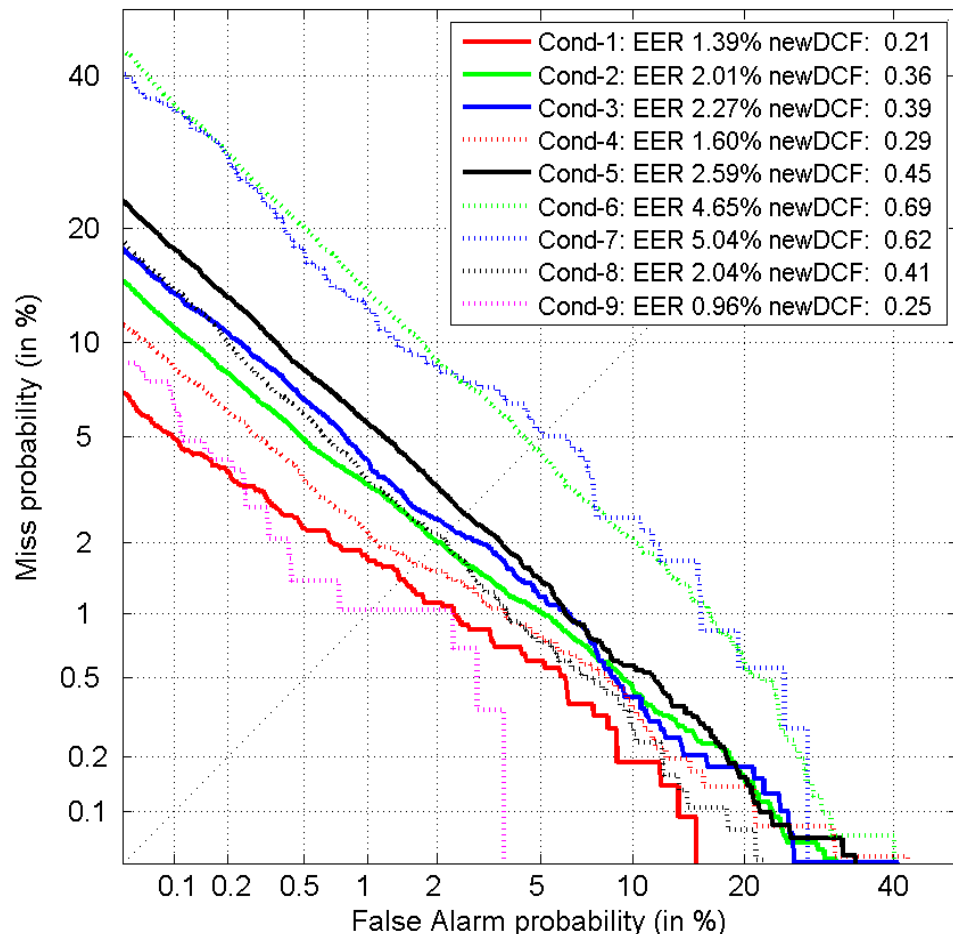


# The end of the story in NIST 2010

PLDA-MFCC



PLDA-LFCC



- This is the beginning of NIST 2011 analysis workshop

# How to do the analysis?

- Unify all the meta-data + VAD + scores:
  - Speaker
  - Segment
  - Trial
- Use Matlab to analyze the data:
  - Scatter plot with 2 million points-> Reboot system
- Need for optimized data visualization tools

# Ask for advice!!!

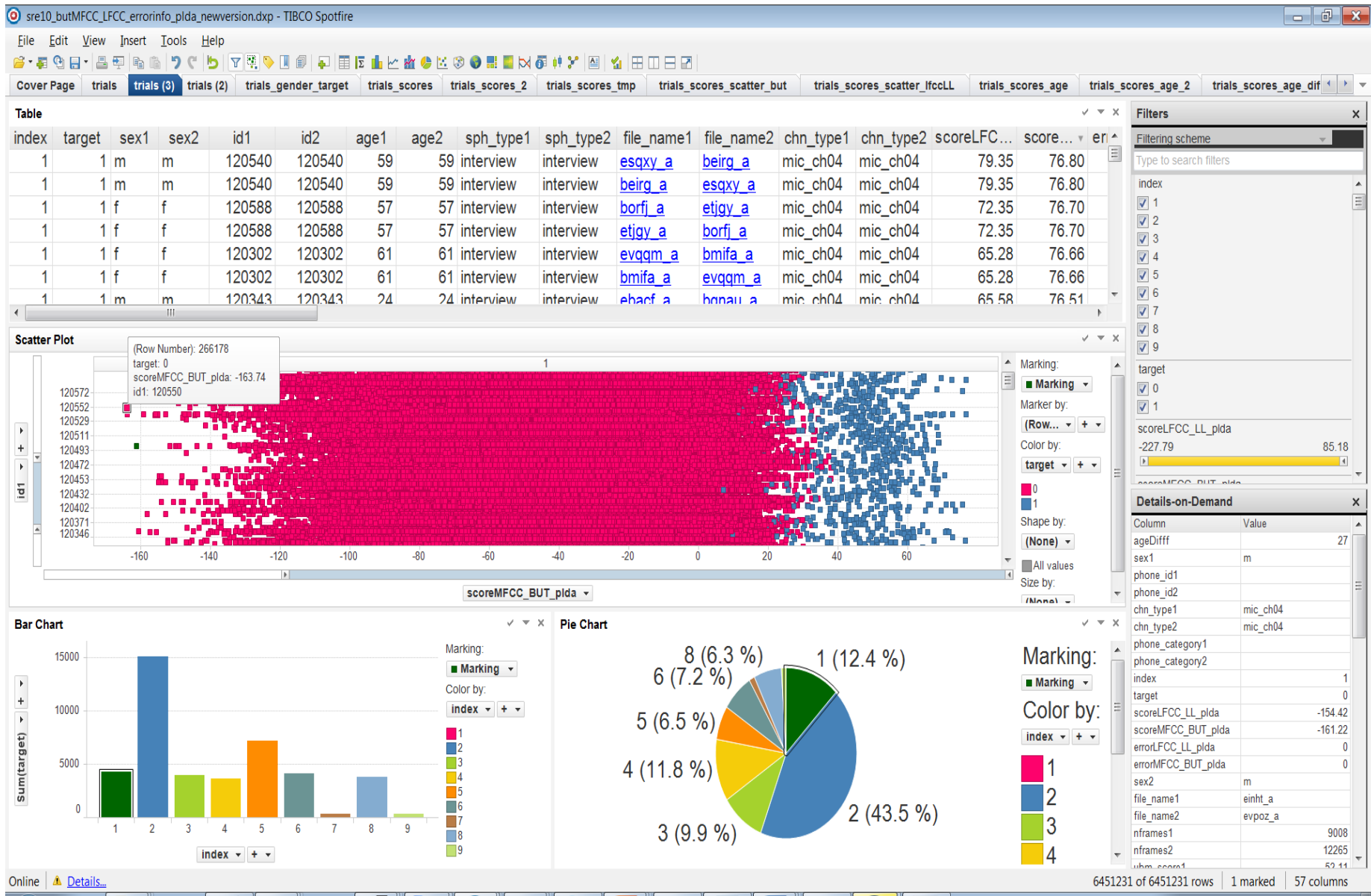
- Talked to Ben Shneiderman:
  - Member of the National Academy of Engineers for contributions to information visualization and knowledge discovery
- Recommended TIBCO Spotfire:
  - Not the only option available: Tableau, Ggplot, TableLens, etc
- Also told me the “mantra” of knowledge discovery through information visualization:

Overview -> Zoom and Filter -> Details on demand

Overview -> Zoom and Filter -> Details on demand

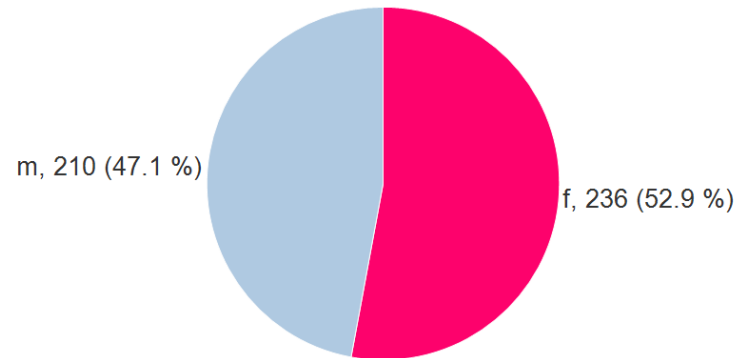
Overview -> Zoom and Filter -> Details on demand

# GUI snapshot of SPOTFIRE

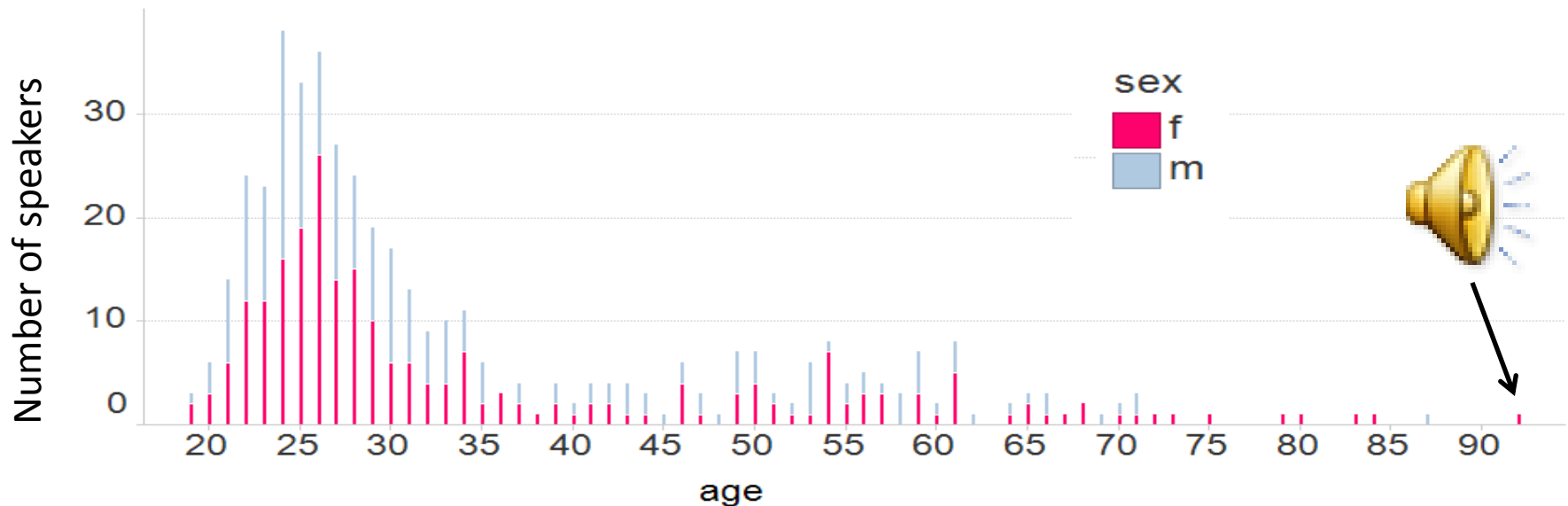


# Overview of speakers: part I

- Total number of speakers: 447

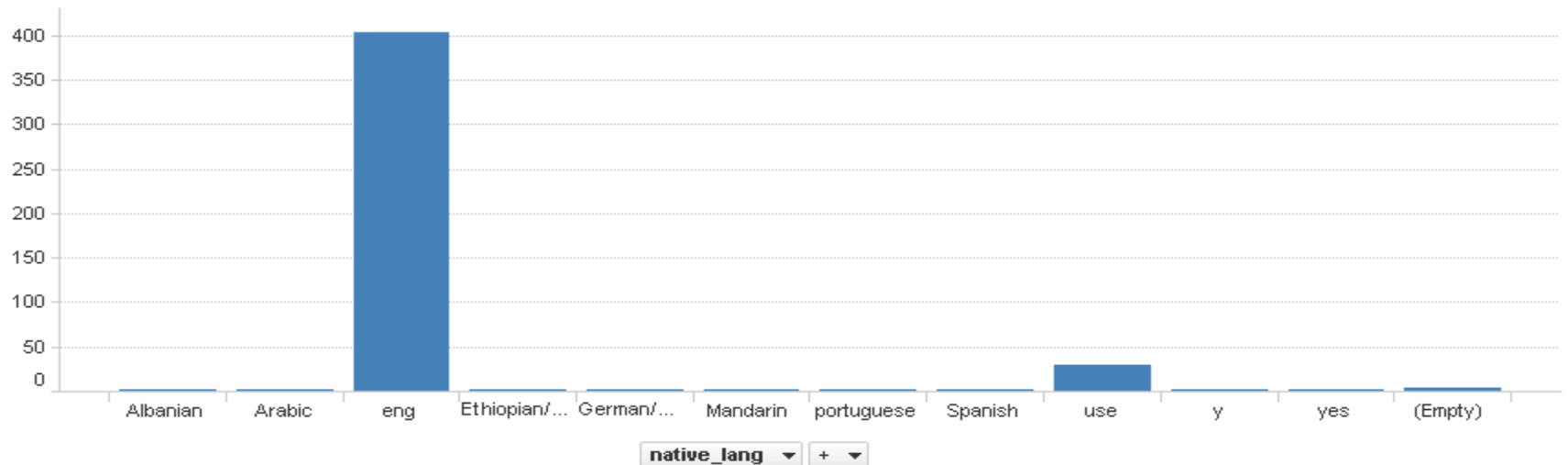


- Age distribution: 19 to 92; Modes -> M:24 and F:26

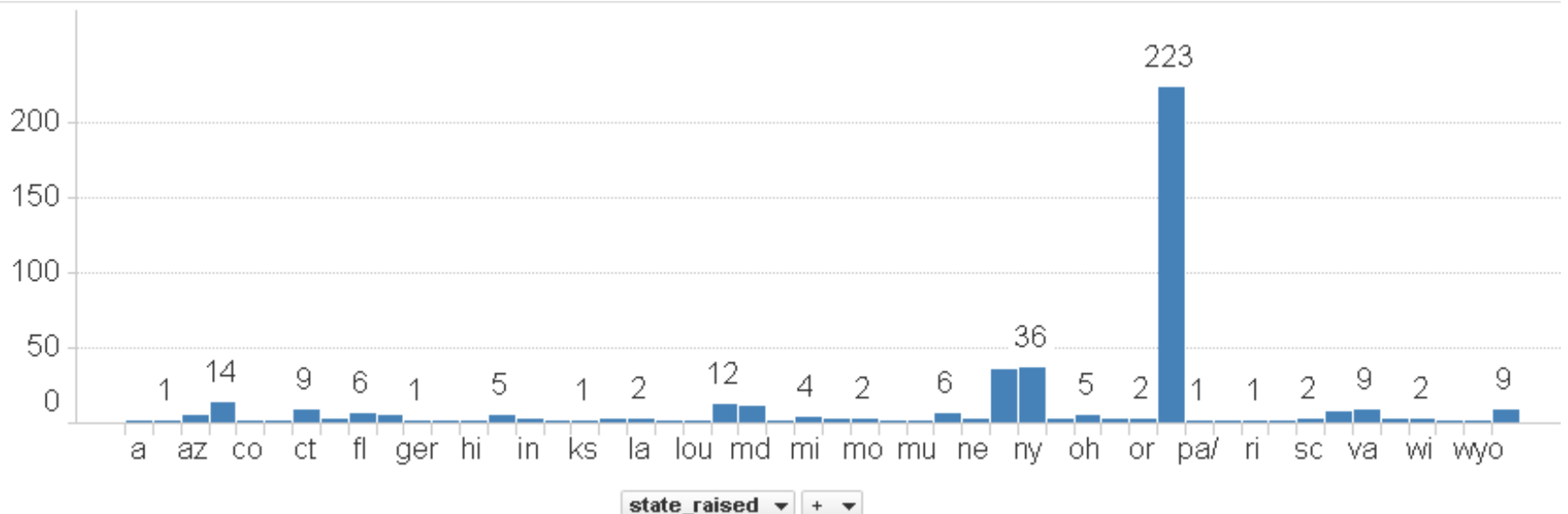


# Overview of speakers: part II

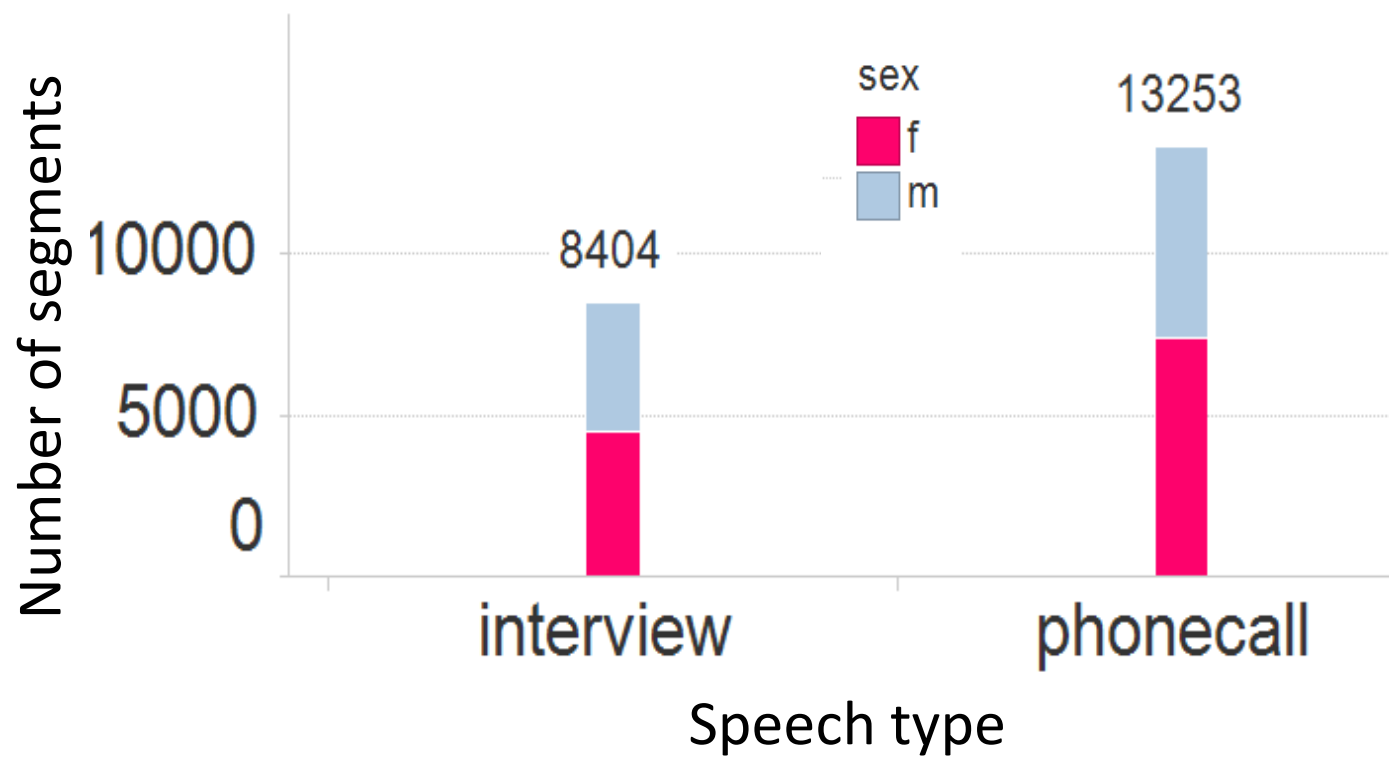
- Native and non-native English speakers:



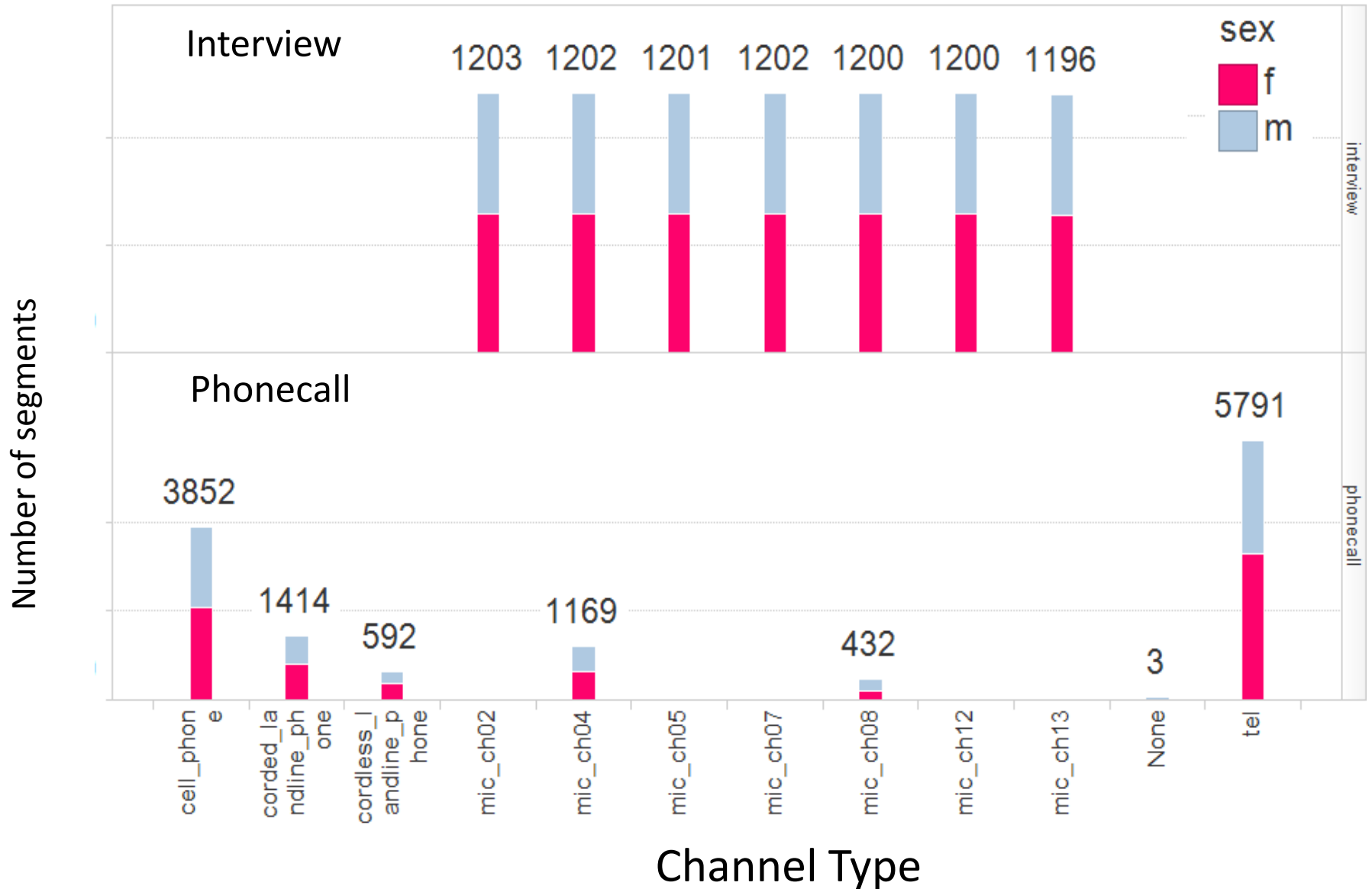
- State raised:



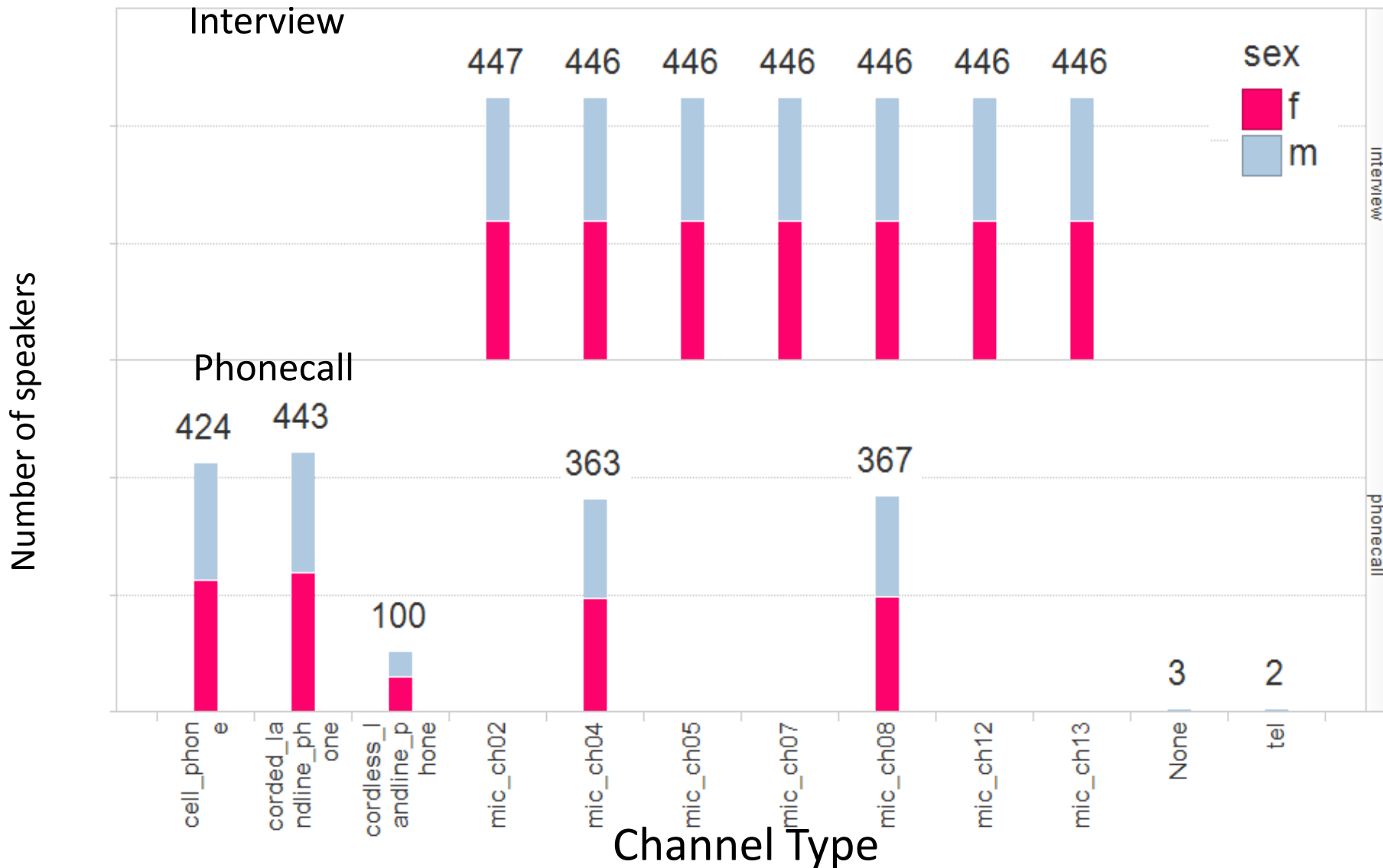
# Overview of speech segments



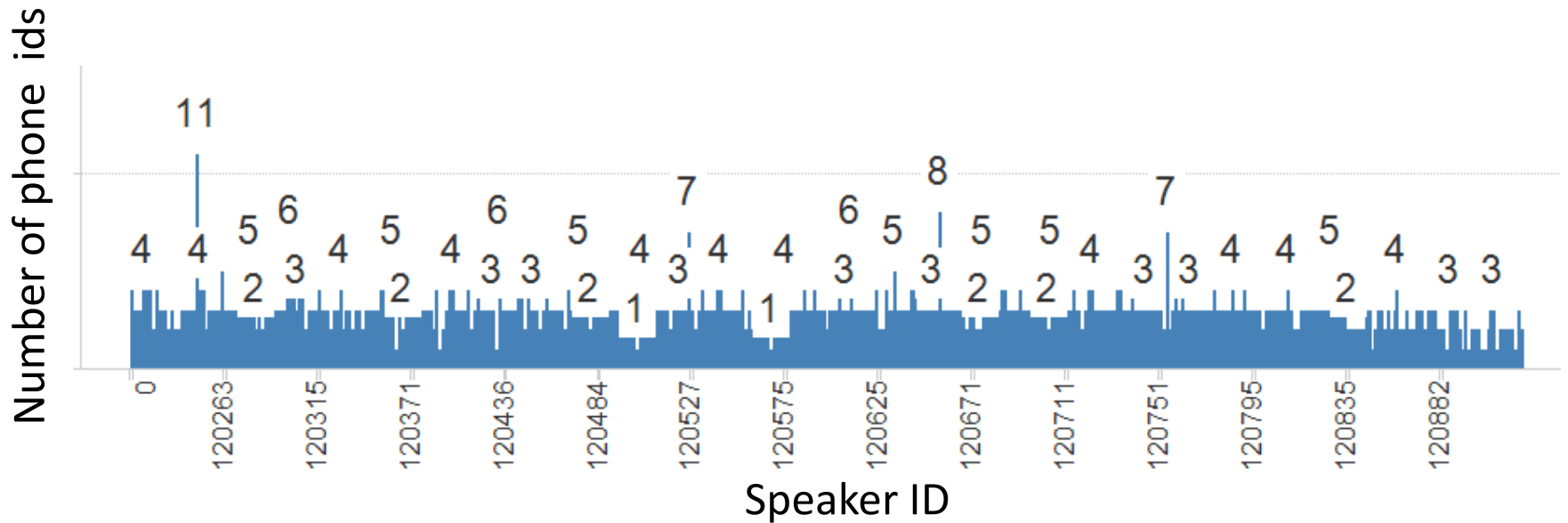
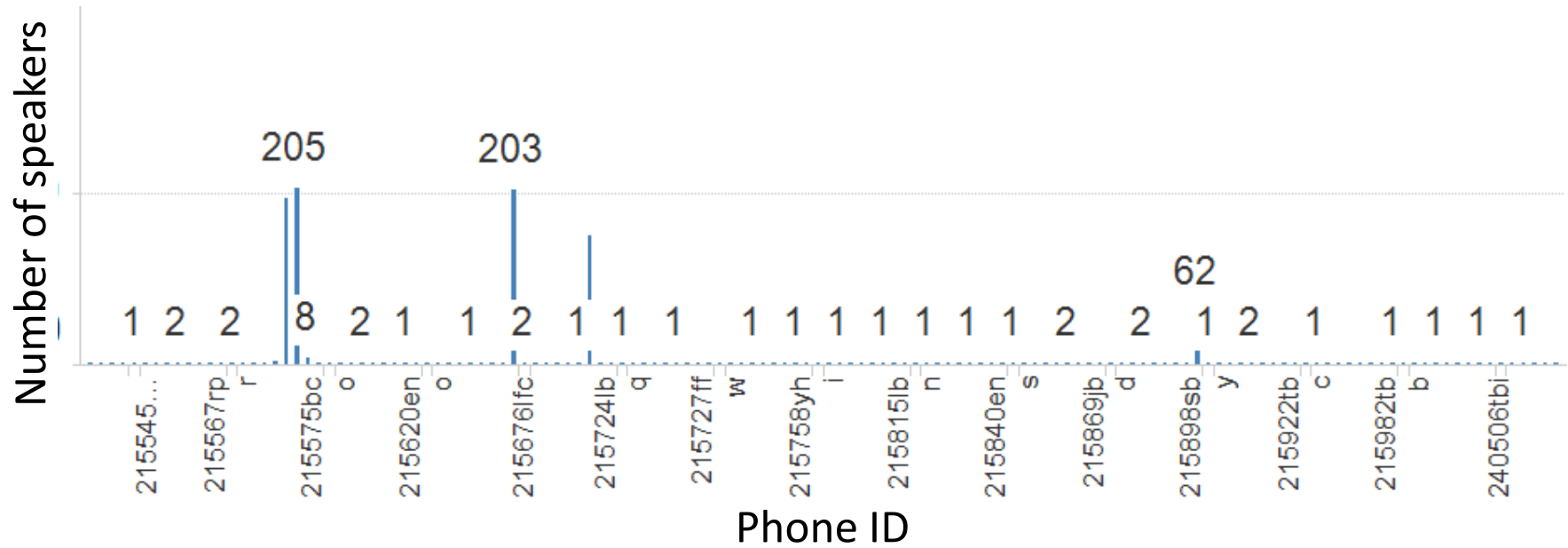
# Segments and channel types



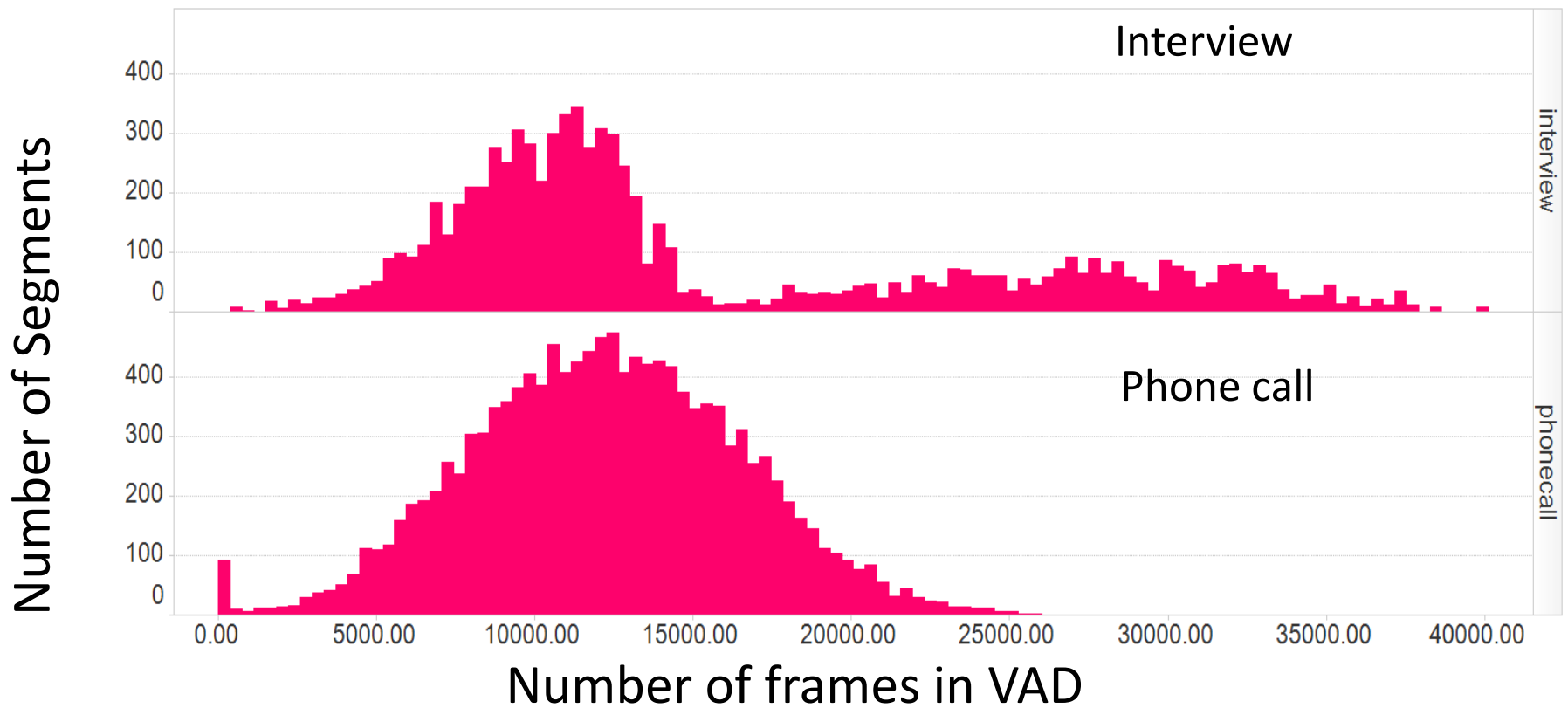
# Speakers and channel types



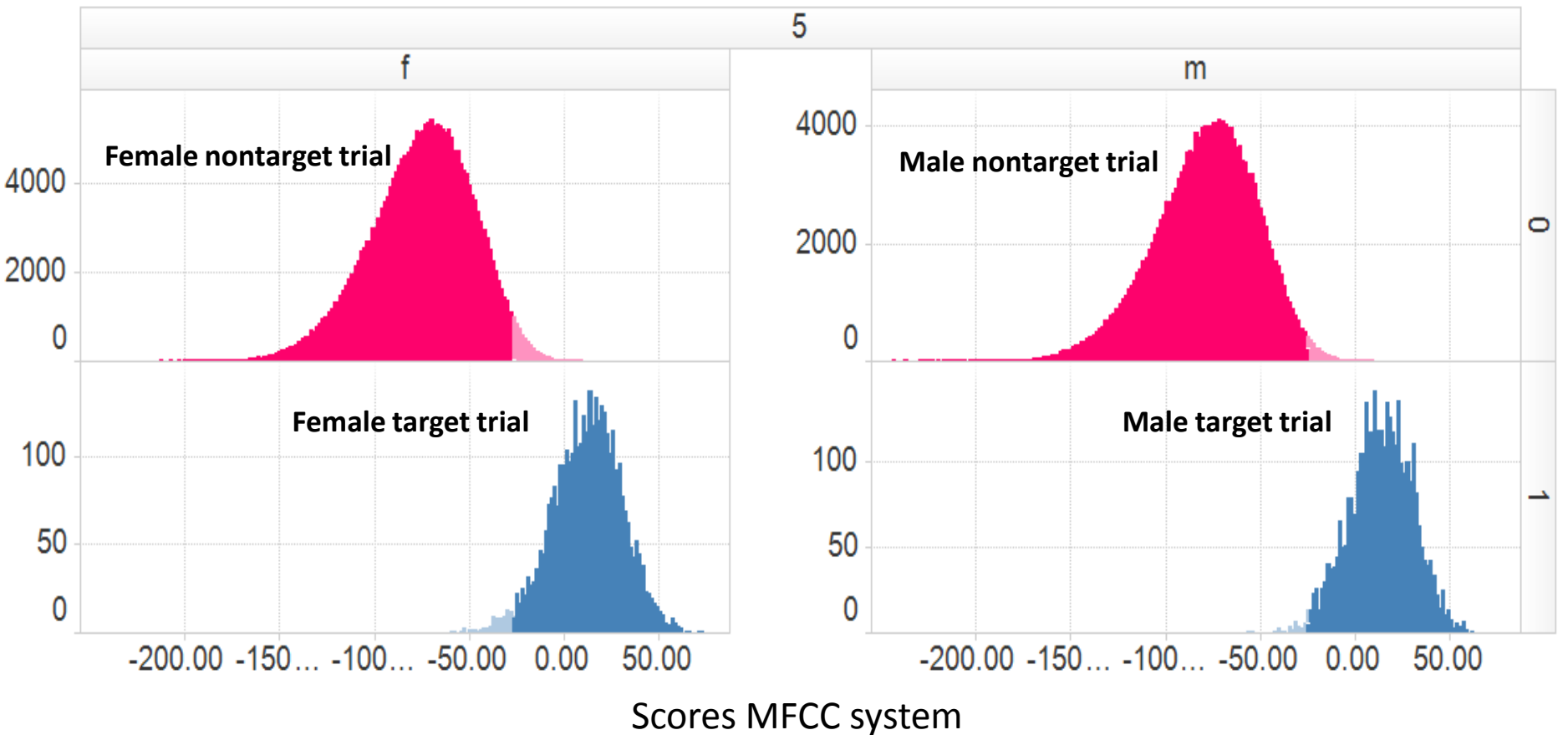
# Speakers and phone IDs



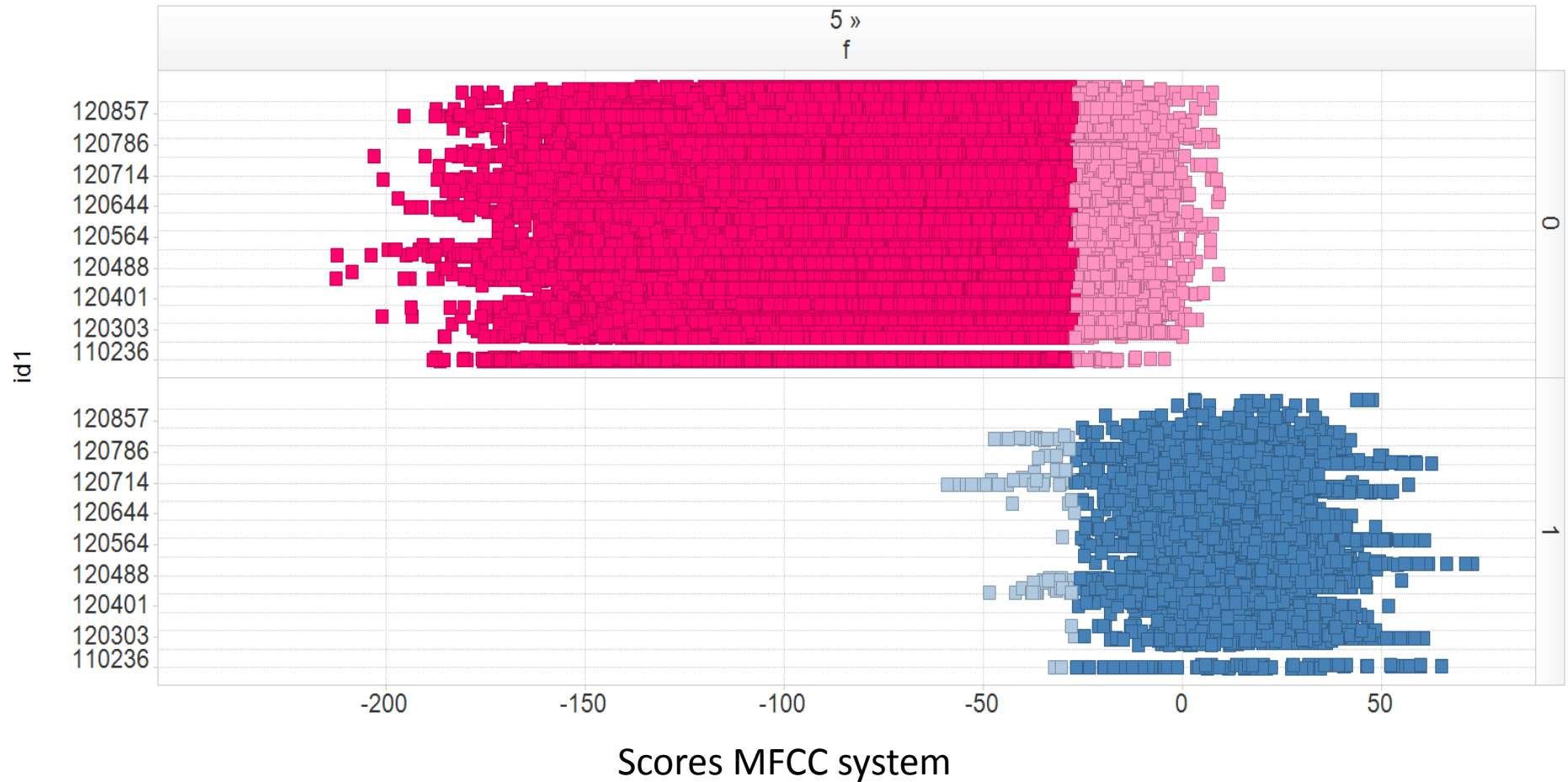
# Segments and VAD



# Histogram of scores



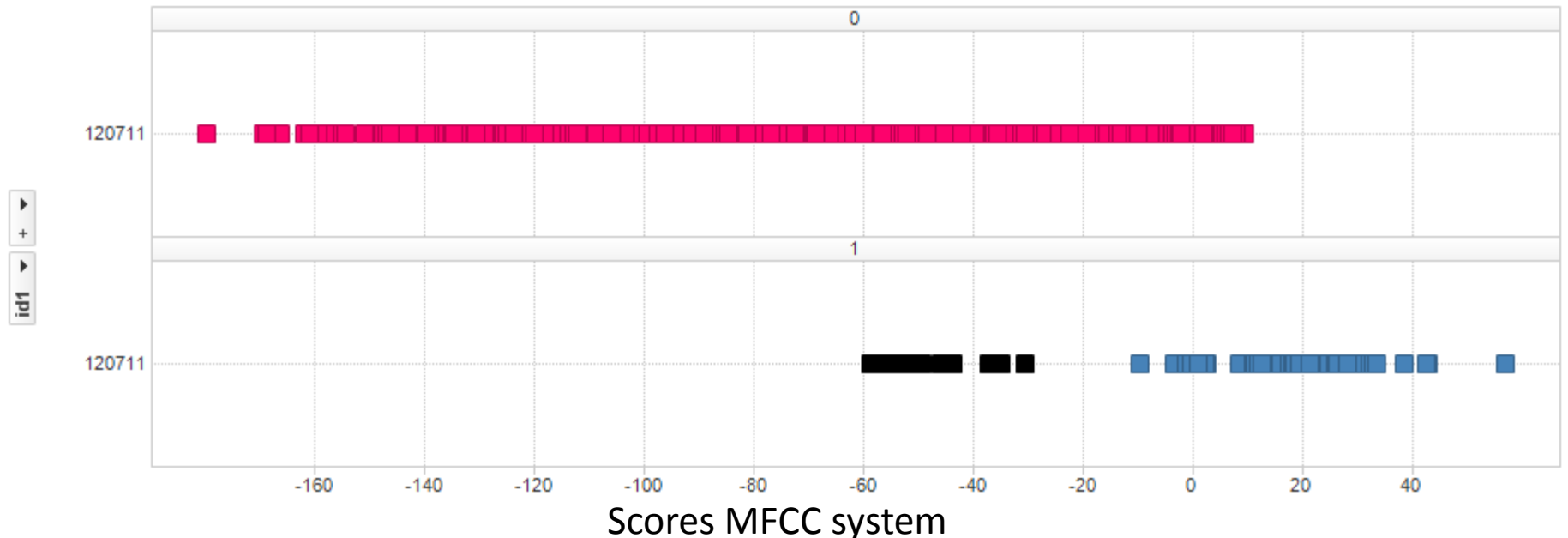
# Scatter plot: speaker-ID vs score



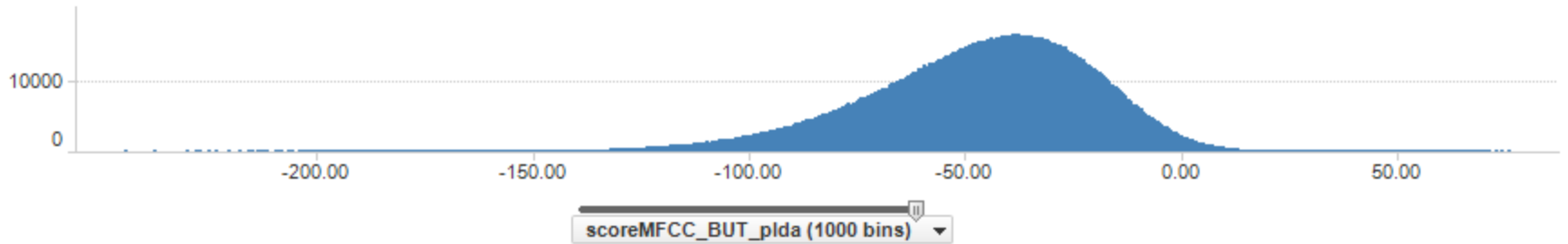
# Zoom on speaker of interest

ind	scor...	sc... ▾	se	s	file_na...	file_n...	chn_type1	chn_type2	nfram...	nfram...	id1	id2
5	-20.75	-30.73	f	f	<a href="#">tndnq_a</a>	<a href="#">tltxj_a</a>	cell_phone	corded_landlin...	13145	10165	120711	120711
5	-24.37	-34.97	f	f	<a href="#">ttefb_a</a>	<a href="#">tltxj_a</a>	cell_phone	corded_landlin...	5832	10165	120711	120711
5	-23.35	-37.28	f	f	<a href="#">txymn_b</a>	<a href="#">tltxj_a</a>	cell_phone	corded_landlin...	9264	10165	120711	120711
5	-28.24	-43.89	f	f	<a href="#">tduvc_a</a>	<a href="#">tltxj_a</a>	cell_phone	corded_landlin...	14519	10165	120711	120711
5	-38.45	-46.10	f	f	<a href="#">tjais_a</a>	<a href="#">tltxj_a</a>	cell_phone	corded_landlin...	10115	10165	120711	120711
5	-33.93	-49.56	f	f	<a href="#">tlxdh_b</a>	<a href="#">tltxj_a</a>	cell_phone	corded_landlin...	16721	10165	120711	120711
5	-43.50	-50.39	f	f	<a href="#">tgokq_a</a>	<a href="#">tltxj_a</a>	cell_phone	corded_landlin...	11582	10165	120711	120711
5	-31.18	-52.21	f	f	<a href="#">tmlmd_a</a>	<a href="#">tltxj_a</a>	cell_phone	corded_landlin...	5885	10165	120711	120711
5	-39.15	-52.76	f	f	<a href="#">tbubs_a</a>	<a href="#">tltxj_a</a>	cell_phone	corded_landlin...	15161	10165	120711	120711
5	-32.81	-53.09	f	f	<a href="#">tsxde_a</a>	<a href="#">tltxj_a</a>	cell_phone	corded_landlin...	14552	10165	120711	120711

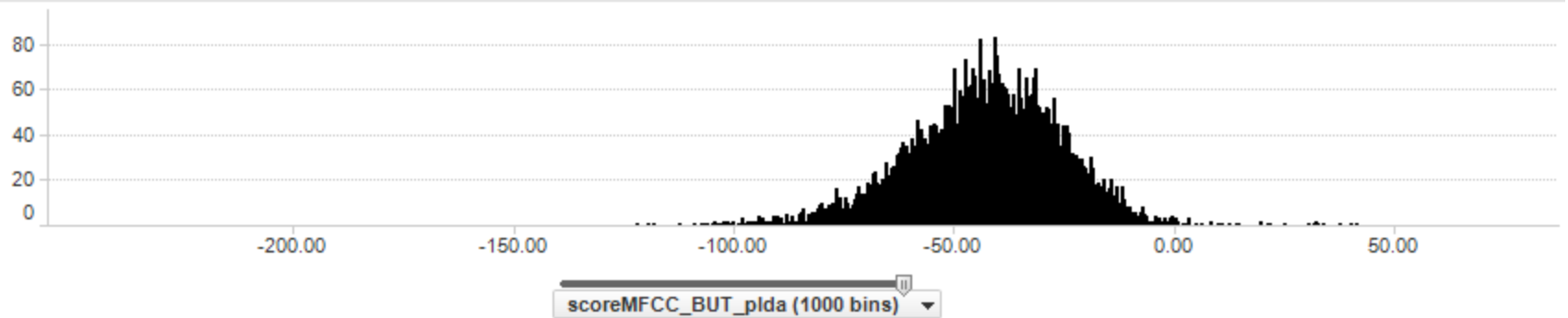
Scatter Plot



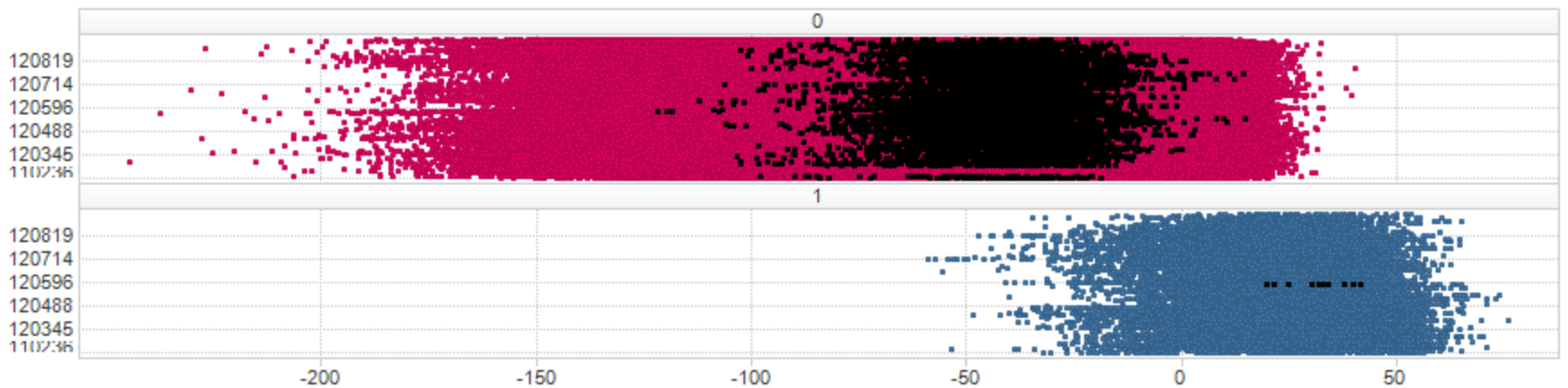
# The case of the 92 year old



art



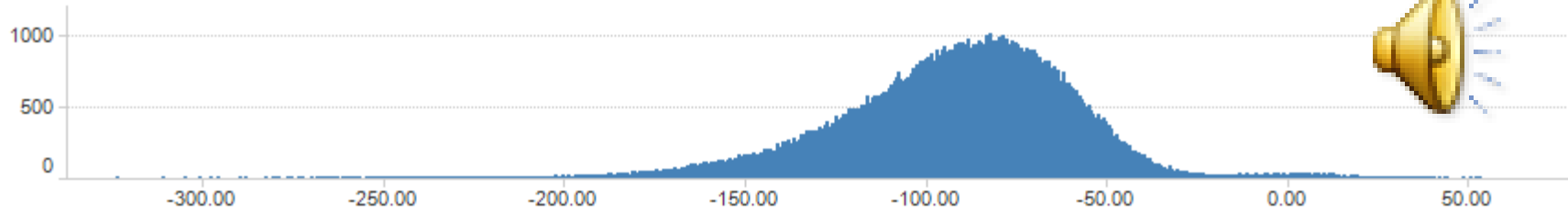
r Plot



# Speaker 120400 identifies himself as a male

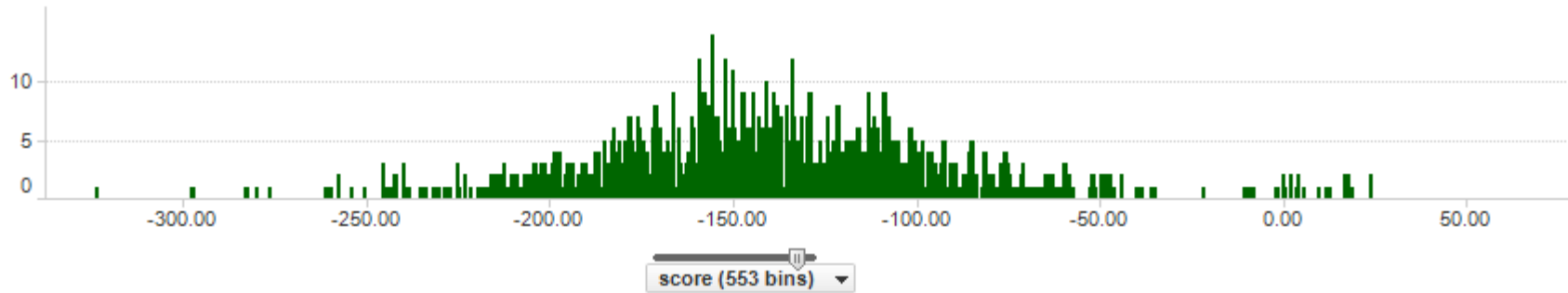


▼  
+  
▼  
(Row C...)



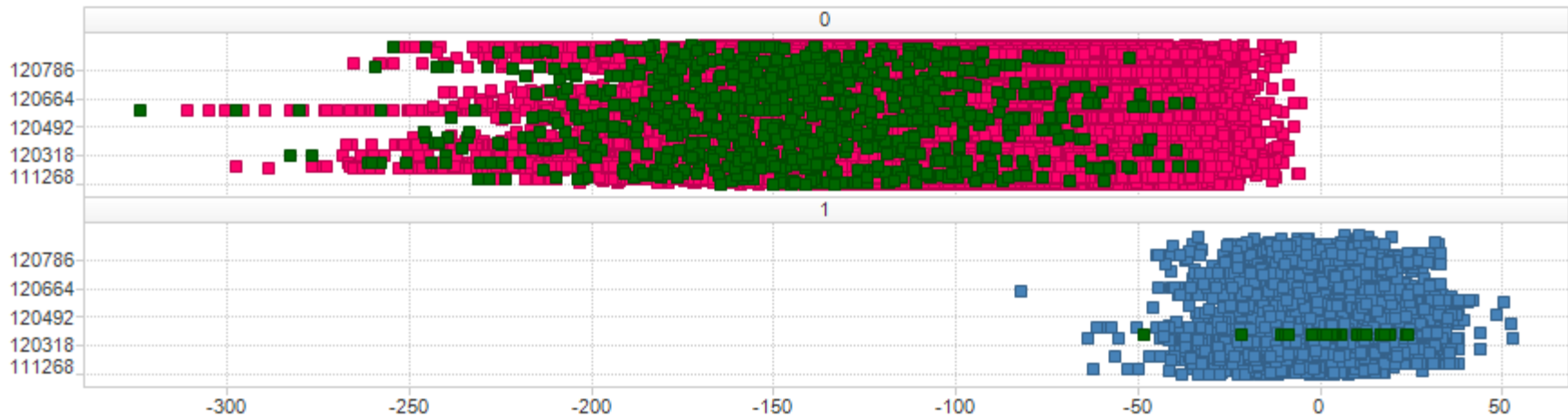
Bar Chart

▼  
+  
▼  
(Row Colu...)



Scatter Plot

▼  
+  
▼  
id1

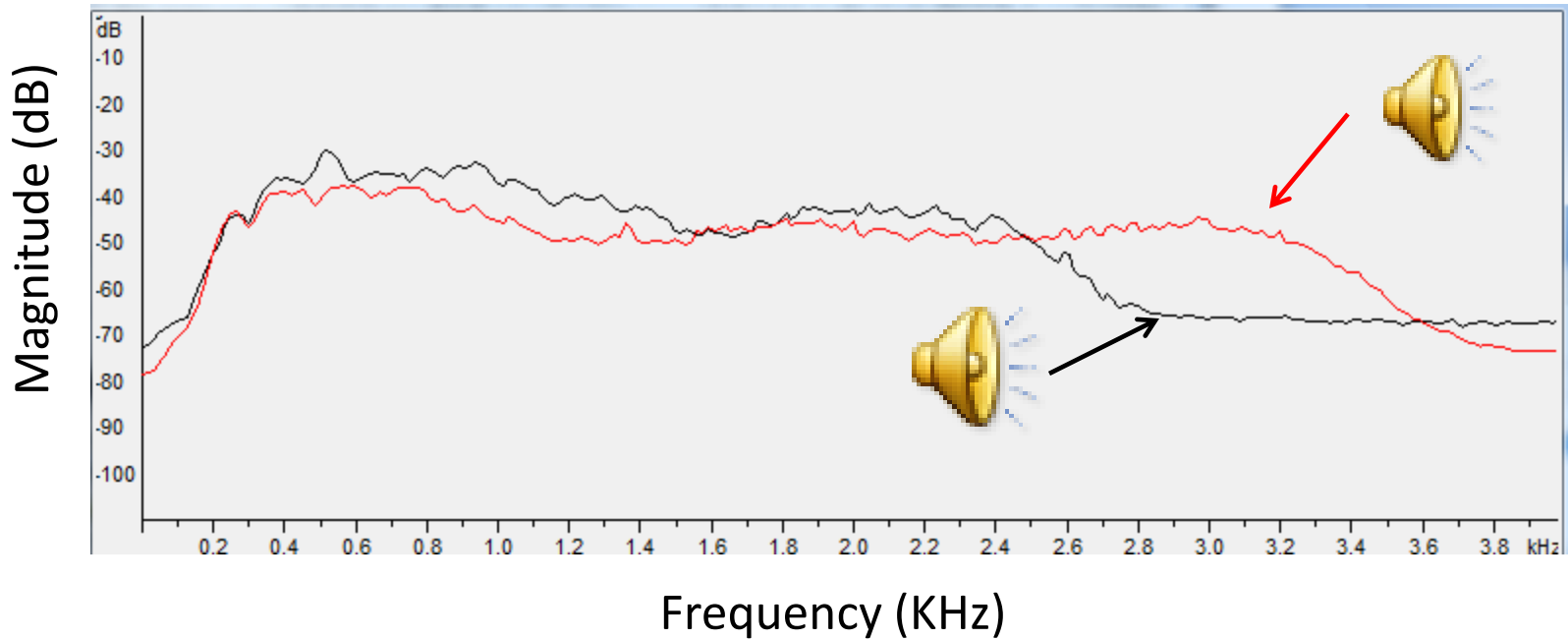


# The case of the very low target scores

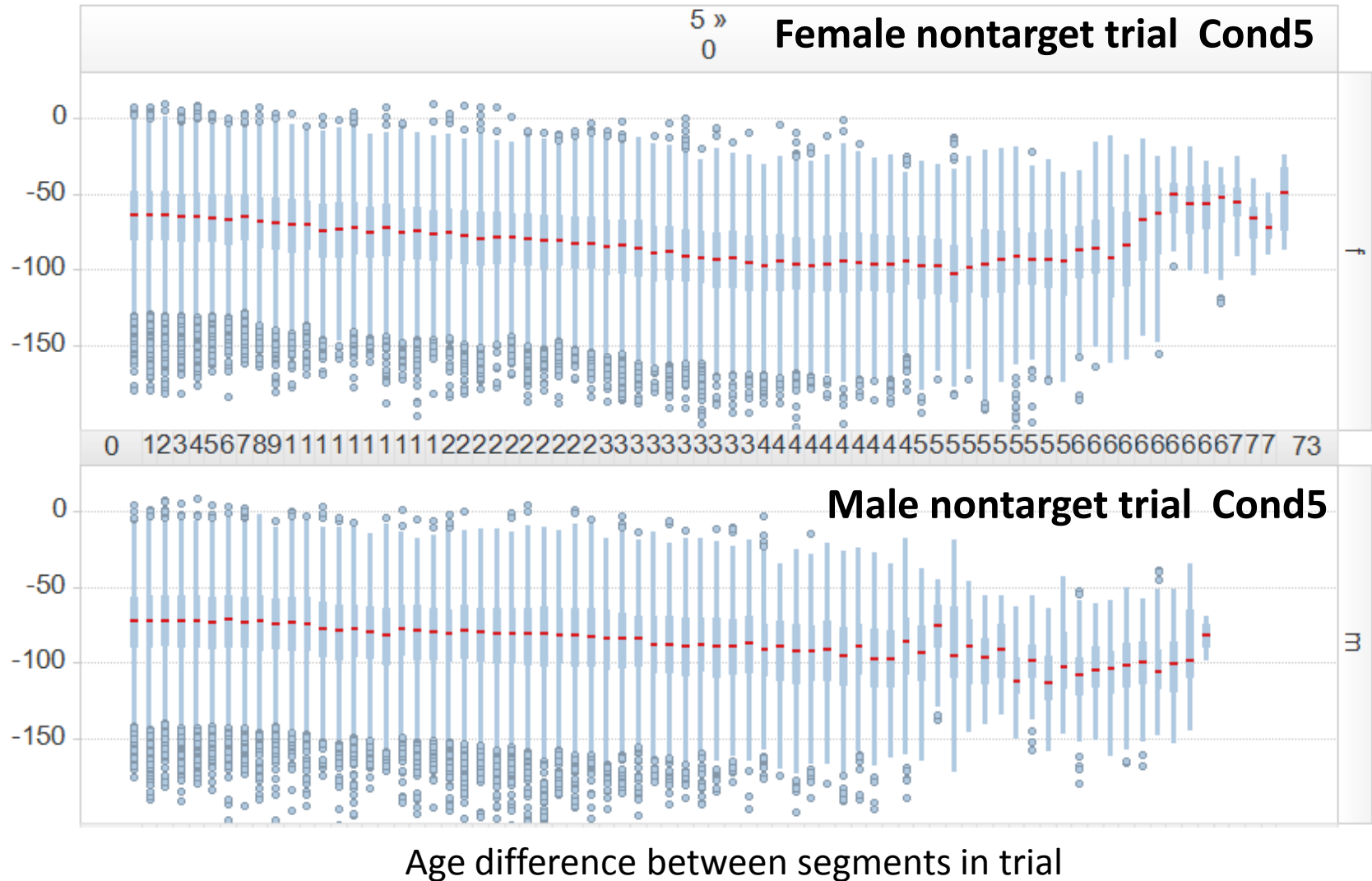


# Strong attenuation above 2.5 KHz

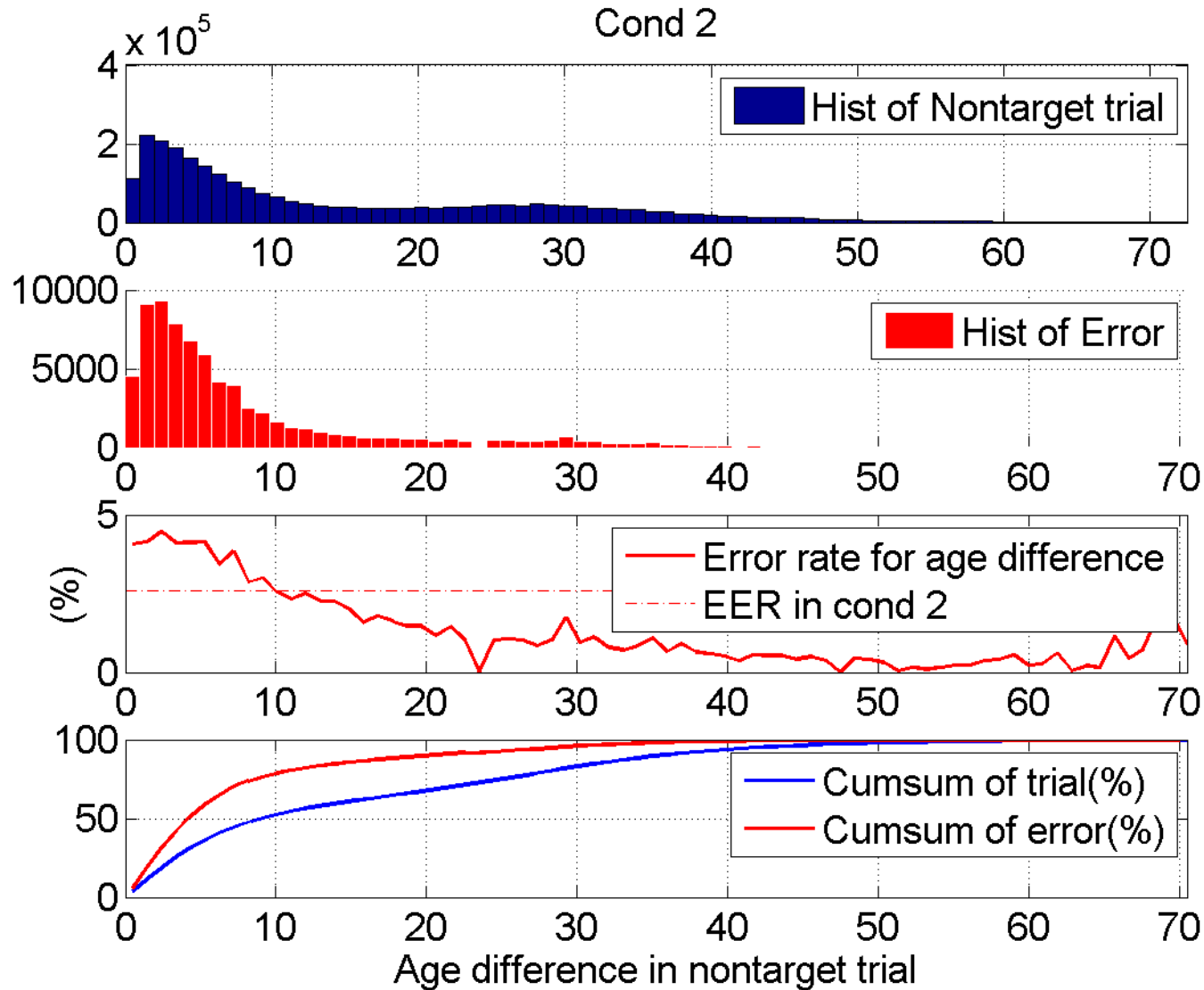
Average short-time spectrum



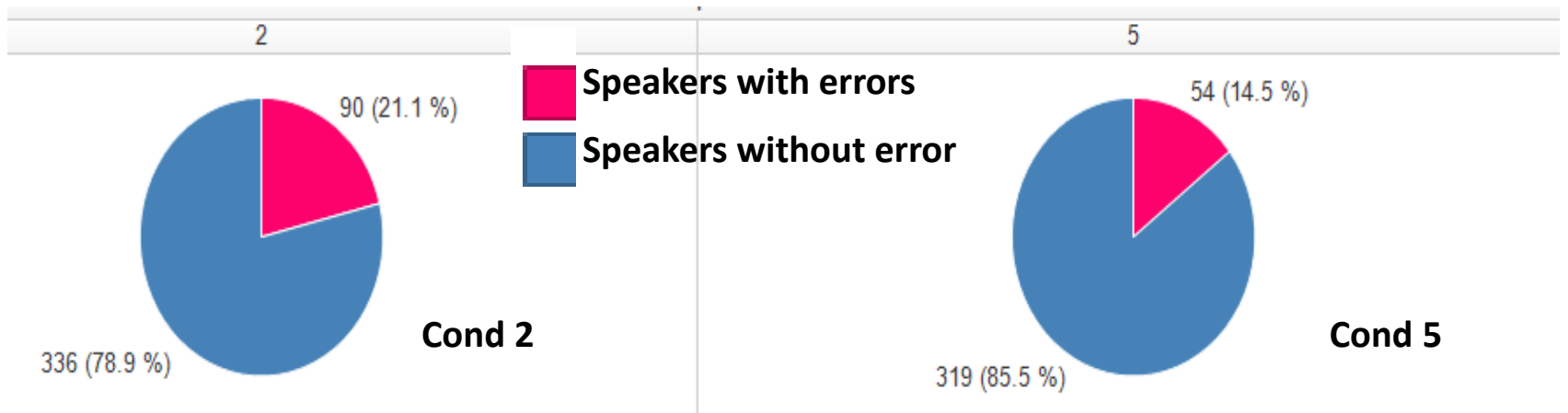
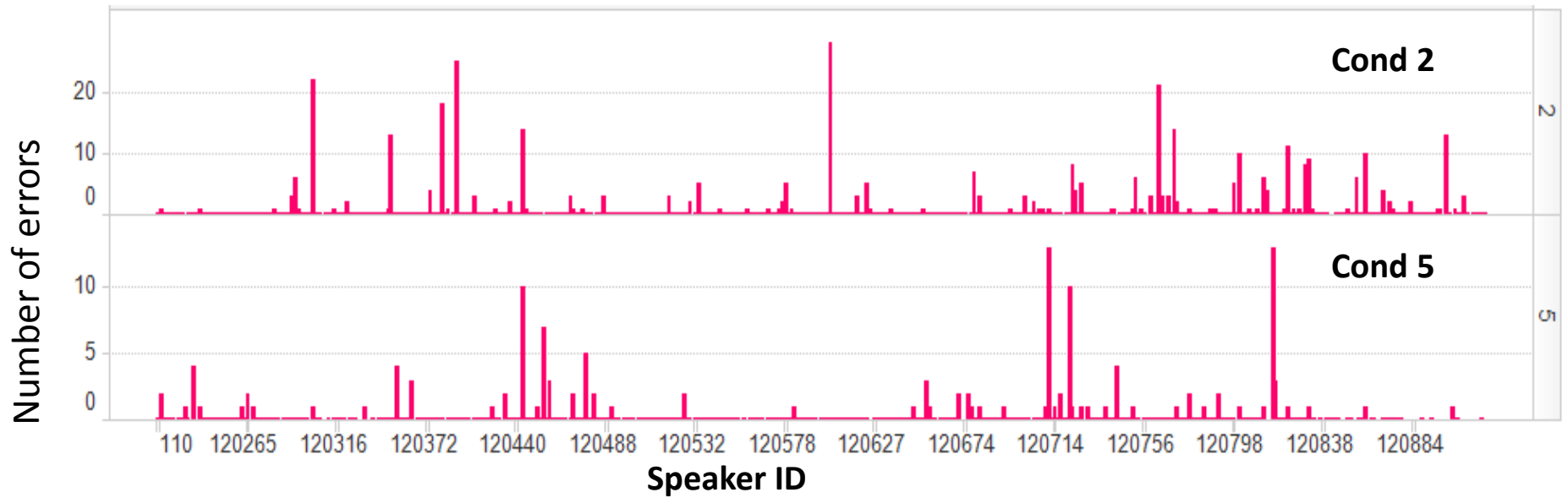
# Influence of age difference in scores



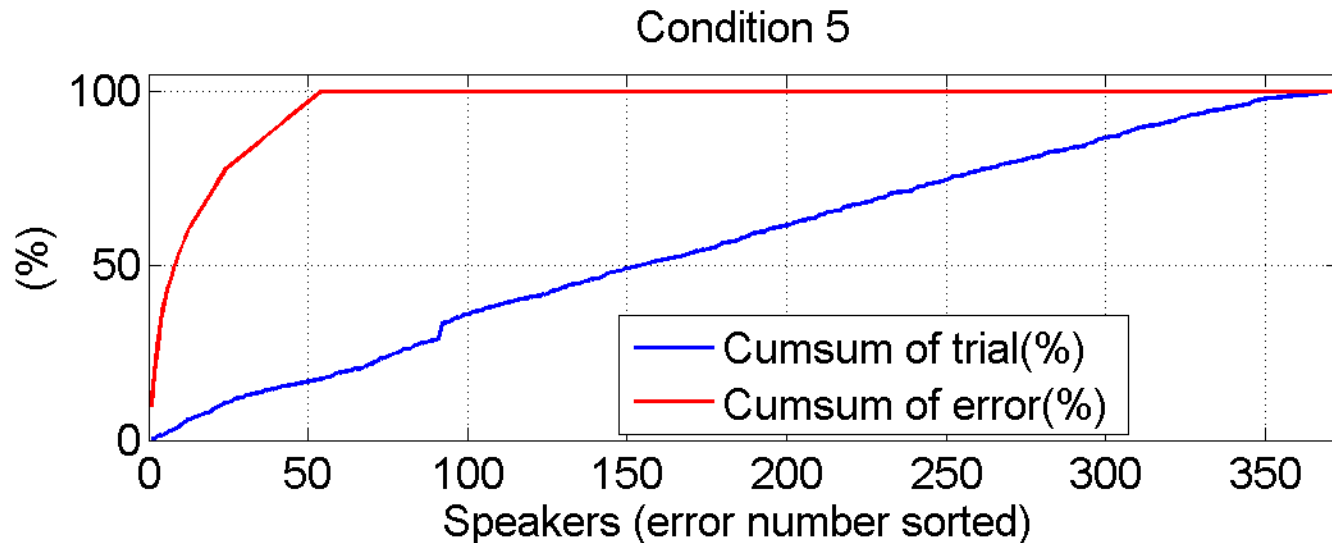
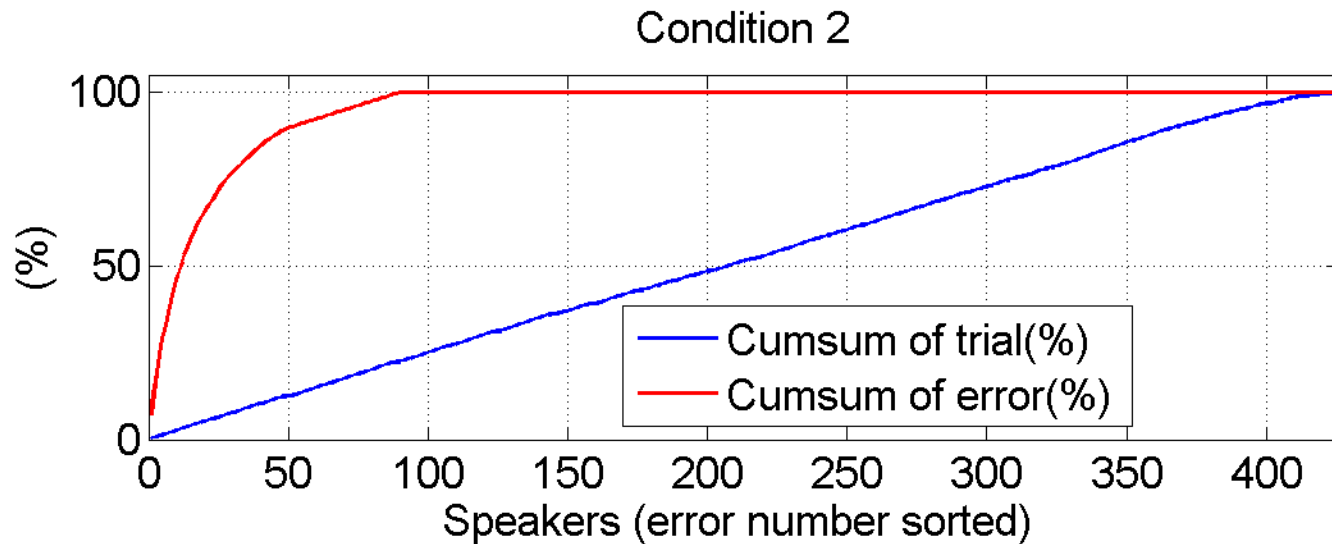
# Influence of age difference in scores



# Miss-detection errors at EER thres



# Miss-detection errors at EER thres



# Conclusions

- Do not attempt to visually handle millions of points in Matlab
- Optimized visualization tools allow you to focus on the analysis rather than on “cosmetics”
- Easy to spot extreme cases
- The PLDA systems seem to behave as expected with respect to average difference in age
- Most of the errors of the system are due to small set of speakers

Thanks!