Consonant confusions in normal and hearing impaired ears
Acoustic Feature Processing

Jont Allen
Andrea Trevino
Christoph Scheidiger
UIUC & Beckman Inst, Urbana IL
Woojae Han, Hallym University, Korea

April 30, 2013

Yogi Berra Quote:

- In theory there is no difference between theory and practice; in practice, there is.

Yogi Berra Quote:

- In theory there is no difference between theory and practice; in practice, there is.

- This form of *paradoxical-contradictions* are known as Yogi-isms

- Intro + Objectives 2 mins
  - Research objectives

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 4 mins $\Sigma 6$
  - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
  - Speech-feature studies ($>$1950)

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 4 mins $\Sigma 6$
  - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
  - Speech-feature studies ($>$1950)
- Methods 8 mins $\Sigma 14$
  - Channel capacity and the Articulation Index
  - Psychophysics of speech/AIgram/3DDS

Outline

- Intro + Objectives — 2 mins
  - Research objectives
- Historical overview — 4 mins $\Sigma 6$
  - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
  - Speech-feature studies (>1950)
- Methods — 8 mins $\Sigma 14$
  - Channel capacity and the Articulation Index
  - Psychophysics of speech/AIgram/3DDS
- Results with NH ears — 10 mins $\Sigma 24$
  - Binary features; Confusions; Primes and Morphs;

Outline

- Intro + Objectives — 2 mins
  - Research objectives
- Historical overview — 4 mins $\Sigma 6$
  - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
  - Speech-feature studies (>1950)
- Methods — 8 mins $\Sigma 14$
  - Channel capacity and the Articulation Index
  - Psychophysics of speech/AIgram/3DDS
- Results with NH ears — 10 mins $\Sigma 24$
  - Binary features; Confusions; Primes and Morphs;
- Results with HI ears — 20 mins $\Sigma 44$
  - Individual differences of consonant confusions

Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 4 mins $\Sigma 6$
  - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
  - Speech-feature studies ($>$1950)
- Methods 8 mins $\Sigma 14$
  - Channel capacity and the Articulation Index
  - Psychophysics of speech/AIgram/3DDS
- Results with NH ears 10 mins $\Sigma 24$
  - Binary features; Confusions; Primes and Morphs;
- Results with HI ears 20 mins $\Sigma 44$
  - Individual differences of consonant confusions
- Summary + Conclusions 6 mins $\Sigma 50$

- When You Come to a Fork in the Road,

- When You Come to a Fork in the Road,
  Take It!

1. Objectives of the UIUC HSR Group

1. My Fork: Repeat Miller Nicely (1955) for >100 CV sounds 2005

1. Objectives of the UIUC HSR Group

1. My Fork: Repeat Miller Nicely (1955) for $>$100 CV sounds 2005
2. Identify acoustic cues in $>$100 CV tokens 2007
   - a) Onset burst, b) Frequency edge, c) Duration, d) $F0$ modulation, e) Voicing 2007-11

1. Objectives of the UIUC HSR Group

1. My Fork: Repeat Miller Nicely (1955) for $>$100 CV sounds 2005
2. Identify acoustic cues in $>$100 CV tokens 2007
   - a) Onset burst, b) Frequency edge, c) Duration, d) $F0$ modulation, e) Voicing 2007-11
   - Consonant recognition is binary 2012

1. Objectives of the UIUC HSR Group

1. My Fork: Repeat Miller Nicely (1955) for $>$100 CV sounds 2005
2. Identify acoustic cues in $>$100 CV tokens 2007

- a) Onset burst, b) Frequency edge, c) Duration, d) $F0$ modulation, e) Voicing 2007-11
- Consonant recognition is binary 2012
- Full analysis of the articulation index (AI) 2012

1. Objectives of the UIUC HSR Group

1. My Fork: Repeat Miller Nicely (1955) for $>$100 CV sounds 2005
2. Identify acoustic cues in $>$100 CV tokens 2007
   - a) Onset burst, b) Frequency edge, c) Duration, d) $F0$ modulation, e) Voicing 2007-11
   - Consonant recognition is binary 2012
   - Full analysis of the articulation index (AI) 2012
3. Measure CV confusions in ≈50 *hearing impaired* ears 2009
   - Characterize hearing impaired (HI) CV confusions, as a function of SNR and hearing level HL(f) 2010

1. Objectives of the UIUC HSR Group

1. My Fork: Repeat Miller Nicely (1955) for $>$100 CV sounds 2005
2. Identify acoustic cues in $>$100 CV tokens 2007
   - a) Onset burst, b) Frequency edge, c) Duration, d) $F0$ modulation, e) Voicing 2007-11
   - Consonant recognition is binary 2012
   - Full analysis of the articulation index (AI) 2012
3. Measure CV confusions in ≈50 *hearing impaired* ears 2009
   - Characterize hearing impaired (HI) CV confusions, as a function of SNR and hearing level HL(f) 2010
4. Explain HI ear feature extraction deficiencies, based on *individual-differences* in CV confusions 2012-13

1. Objectives of the UIUC HSR Group

1. My Fork: Repeat Miller Nicely (1955) for $>$100 CV sounds 2005
2. Identify acoustic cues in $>$100 CV tokens 2007
   - a) Onset burst, b) Frequency edge, c) Duration, d) $F0$ modulation, e) Voicing 2007-11
   - Consonant recognition is binary 2012
   - Full analysis of the articulation index (AI) 2012
3. Measure CV confusions in ≈50 *hearing impaired* ears 2009
   - Characterize hearing impaired (HI) CV confusions, as a function of SNR and hearing level HL(f) 2010
4. Explain HI ear feature extraction deficiencies, based on *individual-differences* in CV confusions 2012-13
   - Hypothesis: HI Consonant discrimination in noise is due to:

     $\Rightarrow$ Cochlear Dead regions?

     $\Rightarrow$ Poor acoustic time/freq edge detection?

     $\Rightarrow$ Auditory plasticity?

Yogi Berra Quote:

- If you don’t know where you’re going,

Yogi Berra Quote:

- If you don’t know where you’re going, you might not get there.

Yogi Berra Quote:

- If you don’t know where you’re going, you might not get there.

- You don’t know where you’re going

Yogi Berra Quote:

- If you don't know where you're going,
  you might not get there.

- You don't know where you're going
  if you don't know where you came from.
  –jba

Yogi Berra Quote:

- If you don't know where you're going,
  you might not get there.

- You don't know where you're going
  if you don't know where you came from.
  –jba
  - The history is essential

1. Intro + Objectives — 2 mins
   - Research objectives
2. Historical overview — 4 mins $\Sigma 6$
   - AG Bell (1860), Rayleigh (1910), Shannon (1948)
   - Speech-feature studies (>1950)
3. Methods — 8 mins $\Sigma 14$
   - Channel capacity and the Articulation Index
   - Psychophysics of speech/AIgram/3DDS
4. Results with NH ears — 10 mins $\Sigma 24$
   - Binary features; Confusions; Primes and Morphs;
5. Results with HI ears — 20 mins $\Sigma 44$
   - Individual differences of consonant confusions
6. Summary + Conclusions — 6 mins $\Sigma 50$

- Lord Rayleigh's 1908 and George Campbell's 1910
  - Formal telephone (i.e., AG Bell's) speech studies

- Lord Rayleigh's 1908 and George Campbell's 1910
  - Formal telephone (i.e., AG Bell's) speech studies
- Harvey Fletcher's 1921 Articulation Index AI
  - First publish AI by French and Steinberg 1947
    - Accurate predictions of nonsense syllable scores

- Lord Rayleigh's 1908 and George Campbell's 1910
  - Formal telephone (i.e., AG Bell's) speech studies
- Harvey Fletcher's 1921 Articulation Index AI
  - First publish AI by French and Steinberg 1947
    - Accurate predictions of nonsense syllable scores
- Shannon The theory of Information 1948+
  - G.A. Miller, Heise and Lichten Entropy $\mathcal{H}$ 1951
  - G.A. Miller & Nicely CM $P_{h|s}(SNR)$ 1955

- Lord Rayleigh's 1908 and George Campbell's 1910
  - Formal telephone (i.e., AG Bell's) speech studies
- Harvey Fletcher's 1921 Articulation Index AI
  - First publish AI by French and Steinberg 1947
    - Accurate predictions of nonsense syllable scores
- Shannon The theory of Information 1948+
  - G.A. Miller, Heise and Lichten Entropy $\mathcal{H}$ 1951
  - G.A. Miller & Nicely CM $P_{h|s}(SNR)$ 1955
- Context:
  - G.A. Miller 1951 *Language and communication*
  - G.A. Miller 1962 5-word Grammar $\equiv$ 4 dB of SNR
  - Boothroyd JASA 1968; Boothroyd & Nittrouer 1988
  - Bronkhorst et al. JASA 1993

- NH Feature identification 1950-1990

- NH Feature identification 1950-1990
  - 1950-1980: Haskins Labs
  - 1970-1990: MIT, Bell Labs, IU
  - 1980-2011: ASR research at AT&T, IBM, BBN, CMU
  - 2003-2011: NH Confusions in noise

Consonant Feature Studies

- NH Feature identification 1950-1990
  - 1950-1980: Haskins Labs
  - 1970-1990: MIT, Bell Labs, IU
  - 1980-2011: ASR research at AT&T, IBM, BBN, CMU
  - 2003-2011: NH Confusions in noise
- Cochlear modeling 1920-2000
  - 1910-1950: Bell Labs
  - 1960-2010: MIT + Harvard HSBT
  - 1980-2011: NIH funded University research

Consonant Feature Studies

- NH Feature identification 1950-1990
  - 1950-1980: Haskins Labs
  - 1970-1990: MIT, Bell Labs, IU
  - 1980-2011: ASR research at AT&T, IBM, BBN, CMU
  - 2003-2011: NH Confusions in noise
- Cochlear modeling 1920-2000
  - 1910-1950: Bell Labs
  - 1960-2010: MIT + Harvard HSBT
  - 1980-2011: NIH funded University research
- Hearing Impaired studies
  - 2004-2011: Confusion matrices HSR@UIUC

Allen et. al HSR Experiments 2004-2011

Year	Experiment	Student &Allen	Details	Publication
2004	MN04(MN64)	Phatak	16C+4V SWN	JASA (2007)
2005	MN16R	Phatak, Lovitt	MN55R	JASA (2008)
	HIMCL05	Yoon, Phatak	10 HI ears	JASA (2009)
2006	HINALR05	Yoon *et al.*	10 HI ears	JSLR (2012)
	Verification	Regnier	/ta/	JASA (2008)
	CV06-s/w	Phatak/Regnier	8C+9V SWN/WN	
2007	CV06	Pan	CV06	–
	HL07	Li	Hi/Lo pass	JASA (2009)
2008	TR08	Li	Furui86	ASSP (2009)
2009	3DDS	Allen, Li	plosives	TASLP (2011)
	3DDS	Li	plosives	JASA (2010)
	Verification	Abhinauv	burst mods	JASA (2012)
	Verification	Cvengros	burst mods	Rewrite JASA
	MN64 NZE	Singh	PA07	JASA (2012)
2011	3DDS	Li,Trevino	Fricatives	JASA (2012)
	HINAL11-IV	Han	17 HI ears w NALR	Thesis Ch. 3
2010	HIMCL10-I,II,III	Trevino	17 HI ears @MCL	JASA (2013)

- What Time Is It?

Yogi Berra Quote:

- What Time Is It?
  You Mean Now?

Outline

1. Intro + Objectives — 2 mins
   - Research objectives
2. Historical overview — 4 mins $\Sigma 6$
   - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
   - Speech-feature studies (>1950)
   - Speech-feature studies (>1950)
3. Methods — 8 mins $\Sigma 14$
   - Channel capacity and the Articulation Index
   - Psychophysics of speech/AIgram/3DDS
4. Results with NH ears — 10 mins $\Sigma 24$
   - Binary features; Confusions; Primes and Morphs;
5. Results with HI ears — 20 mins $\Sigma 44$
   - Individual differences of consonant confusions

Summary + Conclusions — 6 mins $\Sigma 50$

Methods: The count (confusion) matrix

- Miller-Nicely's 1955 articulation matrix $P_{h|s}(SNR)$, measured at [-18, -12, -6 shown, 0, 6, 12] dB SNR

TABLE III. Confusion matrix for $S/N = -6$ db and frequency response of 200–6500 cps.

STIMULUS	p	t	k	f	θ	s	ʃ	b	d	g	v	ð	z	ʒ	m	n
p	80	43	64	17	14	6	2	1	1		1	1			2	
t	71	84	55	5	9	3	8	1				1	2		2	3
k	66	76	107	12	8	9	4					1			1	
f	18	12	9	175	48	11	1	7	2	1	2	2				
θ	19	17	16	104	64	32	7	5	4	5	6	4	5			
s	8	5	4	23	39	107	45	4	2	3	1	1	3	2		1
ʃ	1	6	3	4	6	29	195		3							1
b	1			5	4	4		136	10	9	47	16	6	1	5	4
d							8	5	80	45	11	20	20	26	1	
g					2			3	63	66	3	19	37	56		3
v				2		2		48	5	5	145	45	12		4	
ð					6			31	6	17	86	58	21	5	6	4
z					1	1	1	7	20	27	16	28	94	44		1
ʒ								1	26	18	3	8	45	129		2
m	1							4			4	1	3		177	46
n					4			1	5	2		7	1	6	47	163

UNVOICED VOICED NASAL

RESPONSE

Methods: The count (confusion) matrix

- Miller-Nicely's 1955 articulation matrix $P_{h|s}(SNR)$, measured at [-18, -12, -6 shown, 0, 6, 12] dB SNR

TABLE III. Confusion matrix for $S/N = -6$ db and frequency response of 200–6500 cps.

STIMULUS	p	t	k	f	θ	s	ʃ	b	d	g	v	ð	z	ʒ	m	n
p	80	43	64	17	14	6	2	1	1		1	1			2	
t	71	84	55	5	9	3	8	1				1	2		2	3
k	66	76	107	12	8	9	4					1			1	
f	18	12	9	175	48	11	1	7	2	1	2	2				
θ	19	17	16	104	64	32	7	5	4	5	6	4	5			
s	8	5	4	23	39	107	45	4	2	3	1	1	3	2		1
ʃ	1	6	3	4	6	29	195		3							1
b	1			5	4	4		136	10	9	47	16	6	1	5	4
d							8	5	80	45	11	20	20	26	1	
g					2			3	63	66	3	19	37	56		3
v				2		2		48	5	5	145	45	12		4	
ð					6			31	6	17	86	58	21	5	6	4
z					1	1	1	7	20	27	16	28	94	44		1
ʒ								1	26	18	3	8	45	129		2
m	1							4			4	1	3		177	46
n					4			1	5	2		7	1	6	47	163

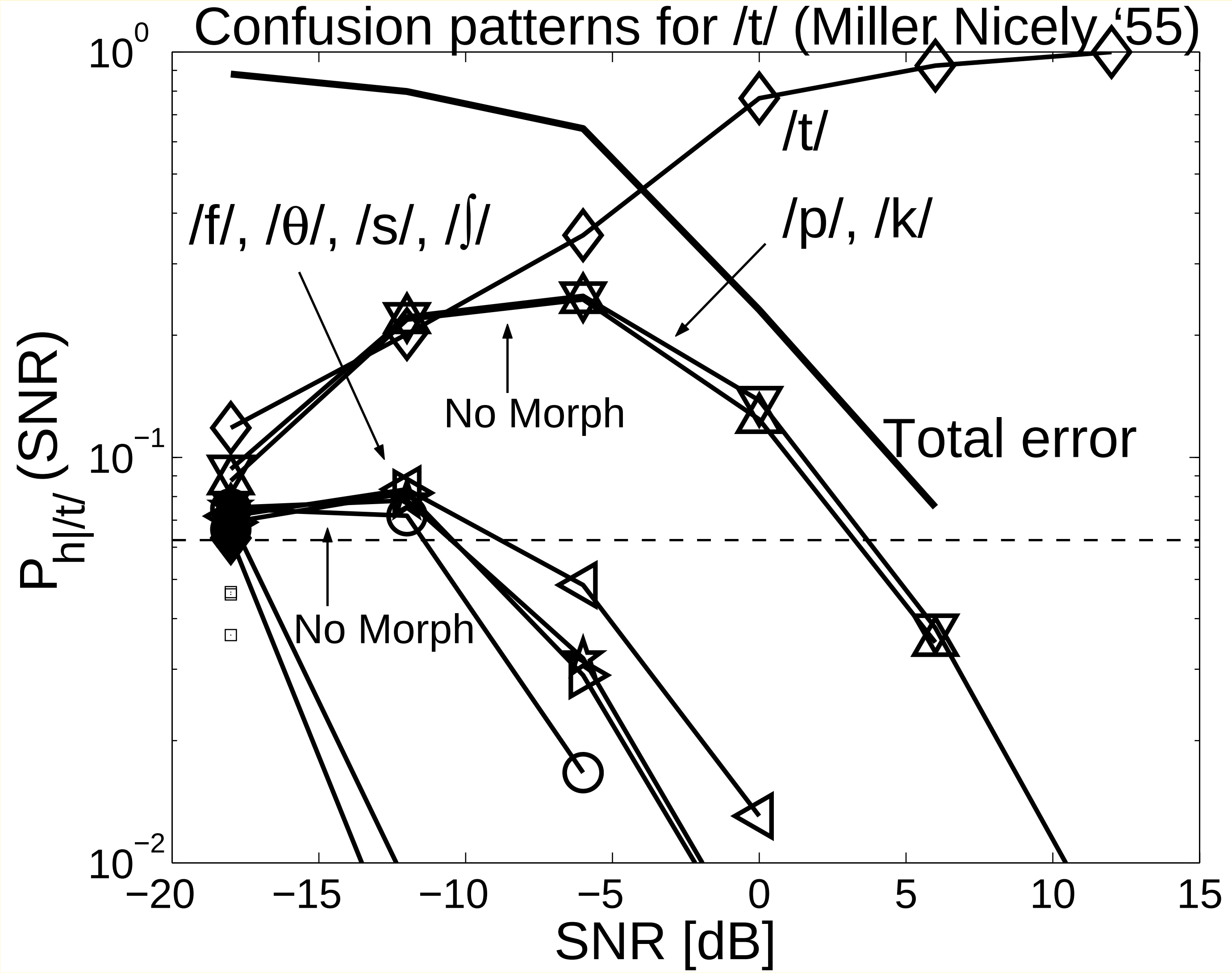
UNVOICED VOICED NASAL

RESPONSE

- Confusion groups $\equiv$ *inhomogeneous confusions*

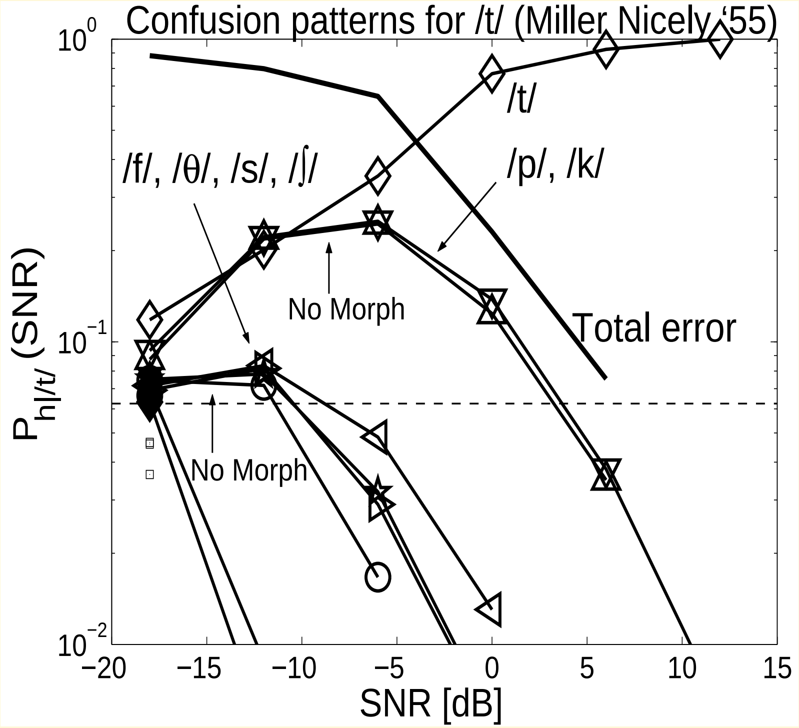
Row of confusion matrix (CM) $P_{h|/t/}$

- This *confusion pattern* characterizes the /t/ row vs SNR



Row of confusion matrix (CM) $P_{h|/t/}$

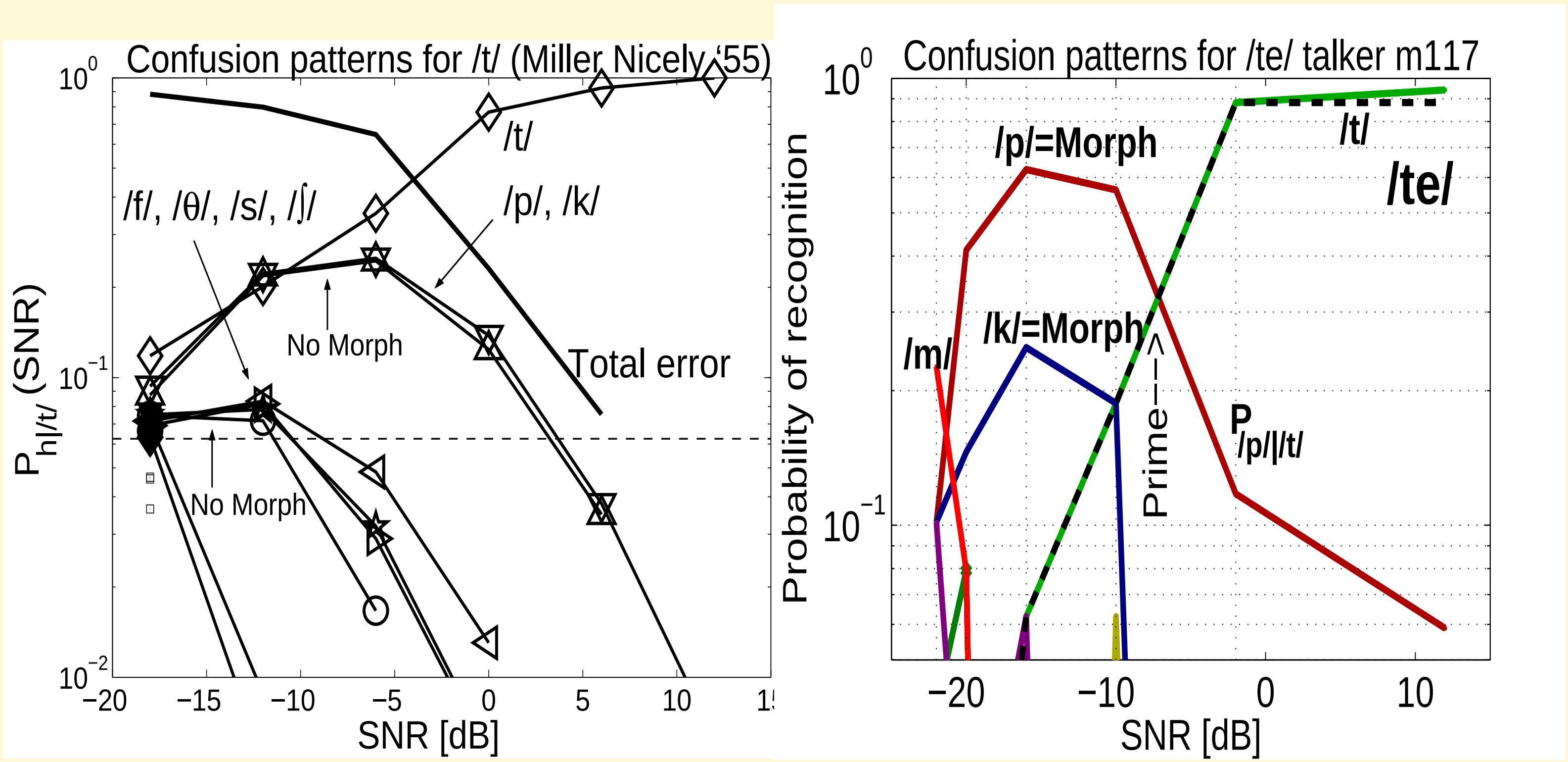
- The SIN_t of averaging across tokens:
  - Token confusions are strongly heterogeneous!
  - Averaging obscures per-token confusions



(a) Average over all /t/s.

Row of confusion matrix (CM) $P_{h|/t/}$

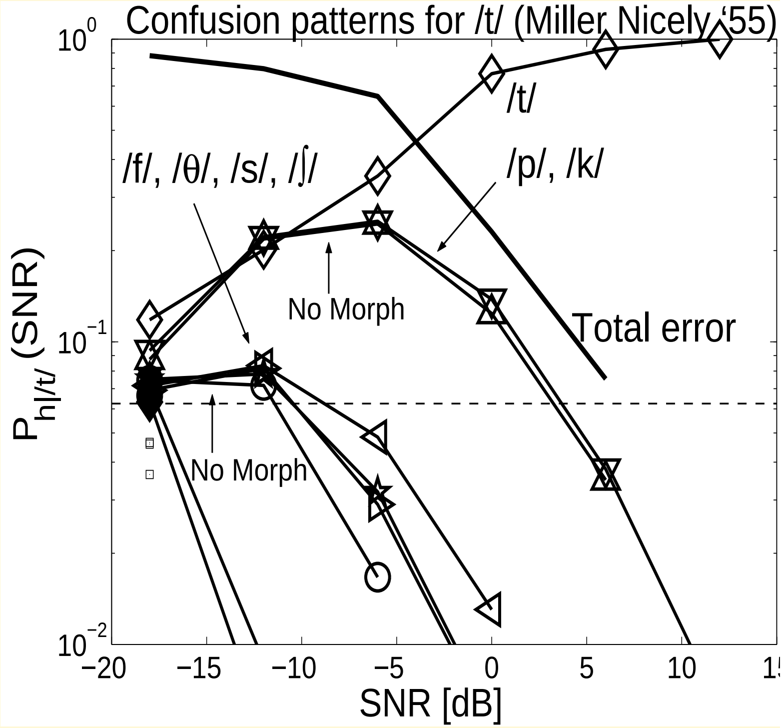
- The SIN_t of averaging across tokens:
  - Token confusions are strongly heterogeneous!
  - Averaging obscures per-token confusions



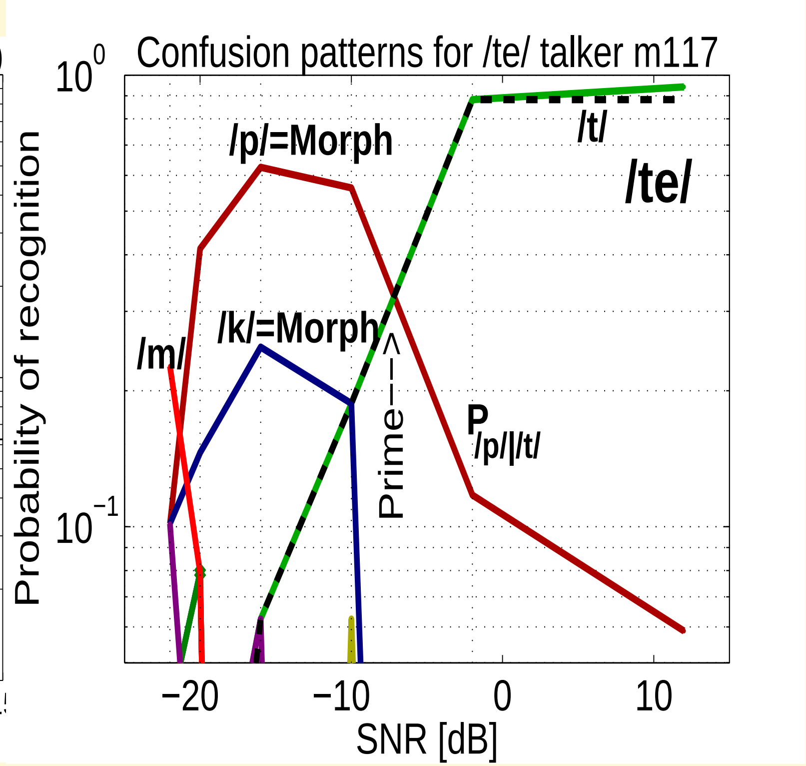
(a) Average over all /t/s.
(b) Talker m117 /te/ $P_{h|/ta/}(SNR)$

Row of confusion matrix (CM) $P_{h|/t/}$

- The SIN_t of averaging across tokens:
  - Token confusions are strongly heterogeneous!
  - Averaging obscures per-token confusions



(a) Average over all /t/s.



(b) Talker m117 /te/ $P_{h|/ta/}(SNR)$

- Never average across tokens!

Methods to Identify Acoustic Features

- Identify the key features in individual CV tokens

Methods to Identify Acoustic Features

- Identify the key features in individual CV tokens
  - -Plosives (e.g., /p, t, k/ and /b, d, g/)
  - -Fricatives (e.g., /θ, ʃ, ʧ, s, h, f/ and /z, ʒ, v, ð/)
  - -With vowels /o, e, ɪ/
    - $\approx$18 talkers and $>$20 listeners
    - Up to 20 trials per consonant per SNR

Methods to Identify Acoustic Features

- Identify the key features in individual CV tokens
  - -Plosives (e.g., /p, t, k/ and /b, d, g/)
  - -Fricatives (e.g., /θ, ʃ, ʧ, s, h, f/ and /z, ʒ, v, ð/)
  - -With vowels /o, e, ı/
    - $\approx$18 talkers and $>$20 listeners
    - Up to 20 trials per consonant per SNR
- Method: 3^d Deep-Search (3DDS) via *truncations:*
  - Time truncation Furui 1986
  - Intensity truncation (i.e., masking)
  - Frequency truncation (High/Low-pass filtering)

Methods to Identify Acoustic Features

- Identify the key features in individual CV tokens
  - -Plosives (e.g., /p, t, k/ and /b, d, g/)
  - -Fricatives (e.g., /θ, ʃ, ʧ, s, h, f/ and /z, ʒ, v, ð/)
  - -With vowels /o, e, ı/
    - $\approx$18 talkers and $>$20 listeners
    - Up to 20 trials per consonant per SNR
- Method: 3^d Deep-Search (3DDS) via *truncations:*
  - Time truncation Furui 1986
  - Intensity truncation (i.e., masking)
  - Frequency truncation (High/Low-pass filtering)
- Methods: Cochlear models & signal processing
  - AIgram Régnier & Allen 2008; Li & Allen 2009,10,11

Yogi Berra Quote:

- Nobody goes there anymore.

Yogi Berra Quote:

- Nobody goes there anymore.
  It's too crowded.

Yogi Berra Quote:

- Nobody goes there anymore.
  It's too crowded.

Do acoustic speech features exist?

- Nobody goes there anymore.
  It's too crowded.

  Do acoustic speech features exist?
  Is there one key acoustic feature per consonant?

- 3^d Deep-Search (3^d-DS) via truncation:

Methods: 3^d Deep Search (3DDS)

- 3^d Deep-Search (3^d-DS) via truncation:
  - Time truncation Furui 1986

Methods: 3^d Deep Search (3DDS)

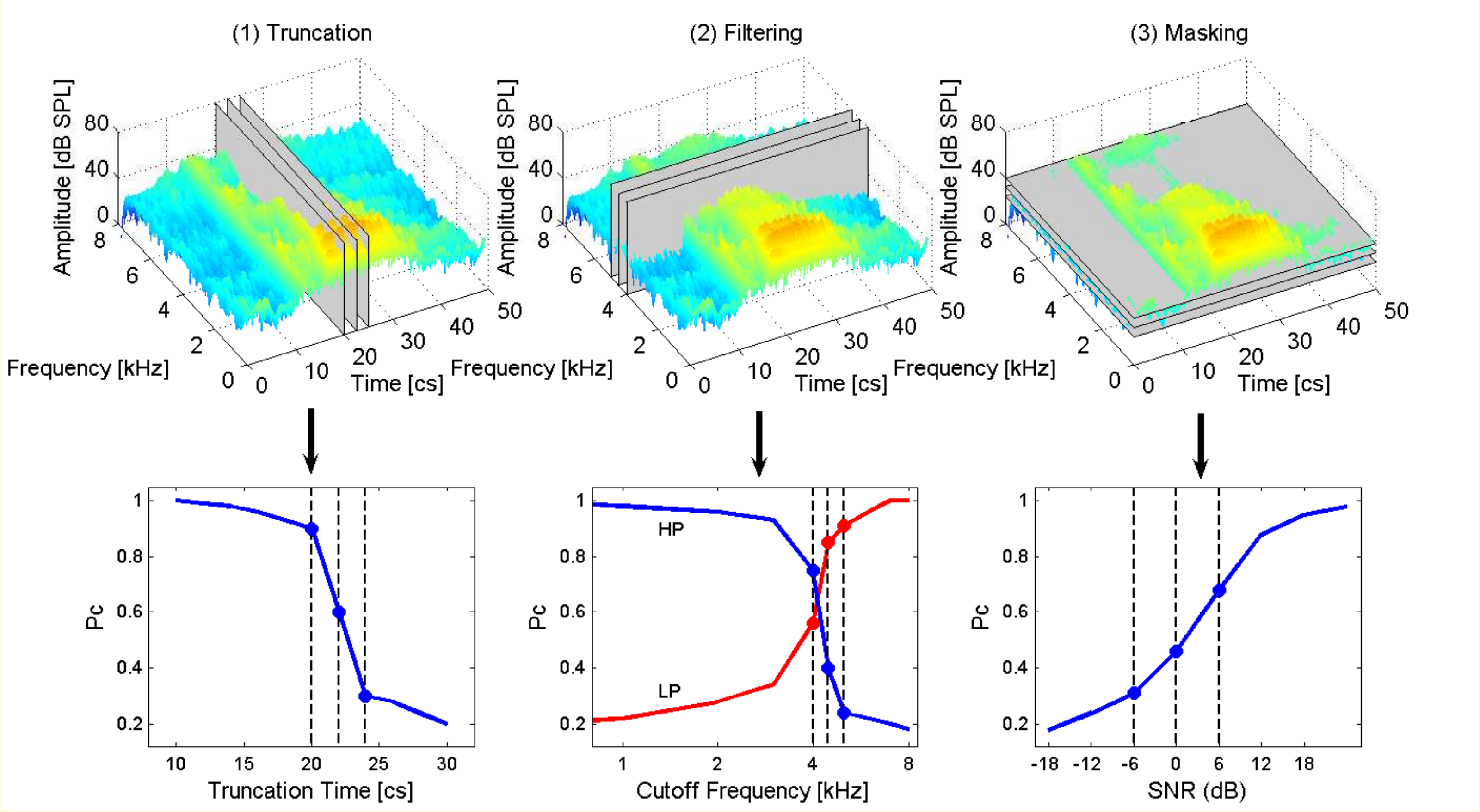
- 3^d Deep-Search (3^d-DS) via truncation:
  - Time truncation Furui 1986
  - Frequency truncation (High/Low-pass filtering)

Methods: 3^d Deep Search (3DDS)

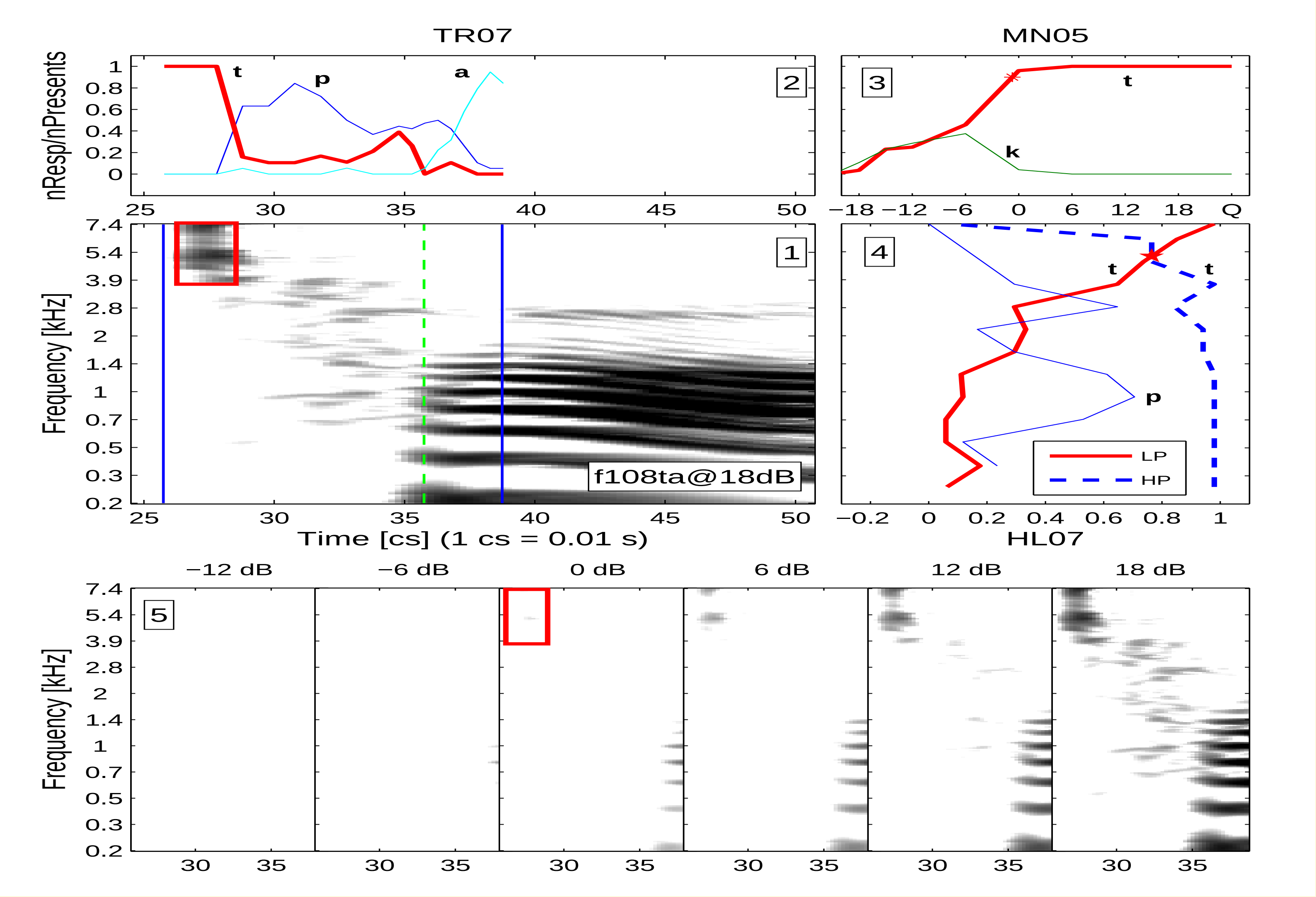
- 3^d Deep-Search (3^d-DS) via truncation:
  - Time truncation Furui 1986
  - Frequency truncation (High/Low-pass filtering)
  - Intensity truncation (i.e., masking)

Methods: 3^d Deep Search (3DDS)

- 3^d Deep-Search (3^d-DS) via truncation:
  - Time truncation Furui 1986
  - Frequency truncation (High/Low-pass filtering)
  - Intensity truncation (i.e., masking)



- Algram: Truncation in Time, Intensity and Frequency

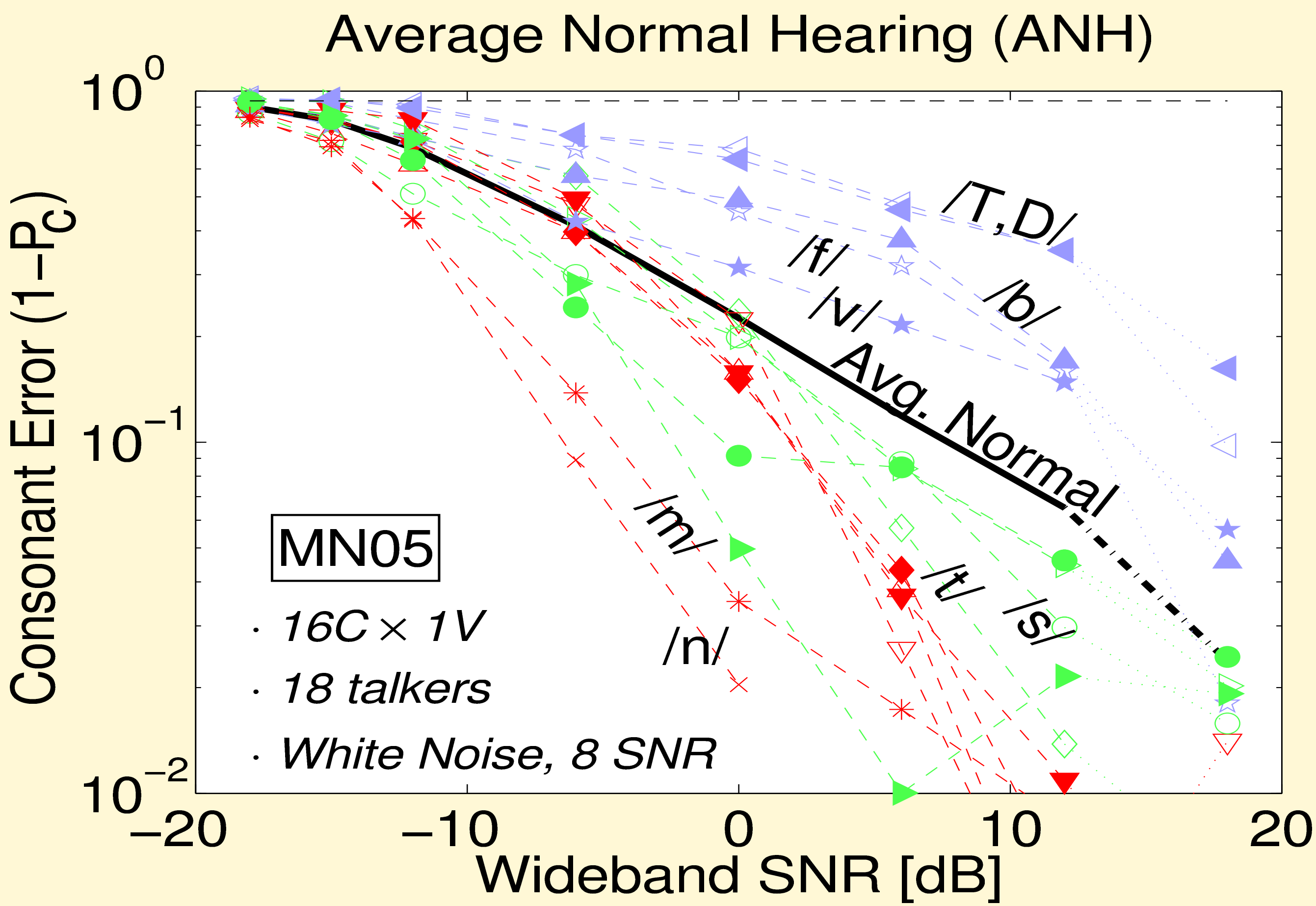


1. Intro + Objectives 2 mins
   - Research objectives
2. Historical overview 4 mins $\Sigma 6$
   - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
   - Speech-feature studies ($>$1950)
3. Methods 8 mins $\Sigma 14$
   - Channel capacity and the Articulation Index
   - Psychophysics of speech/AIgram/3DDS
4. Results with NH ears 10 mins $\Sigma 24$
   - Binary features; Confusions; Primes and Morphs;
5. Results with HI ears 20 mins $\Sigma 46$
   - Individual differences of consonant confusions
6. Summary + Conclusions 6 mins $\Sigma 50$

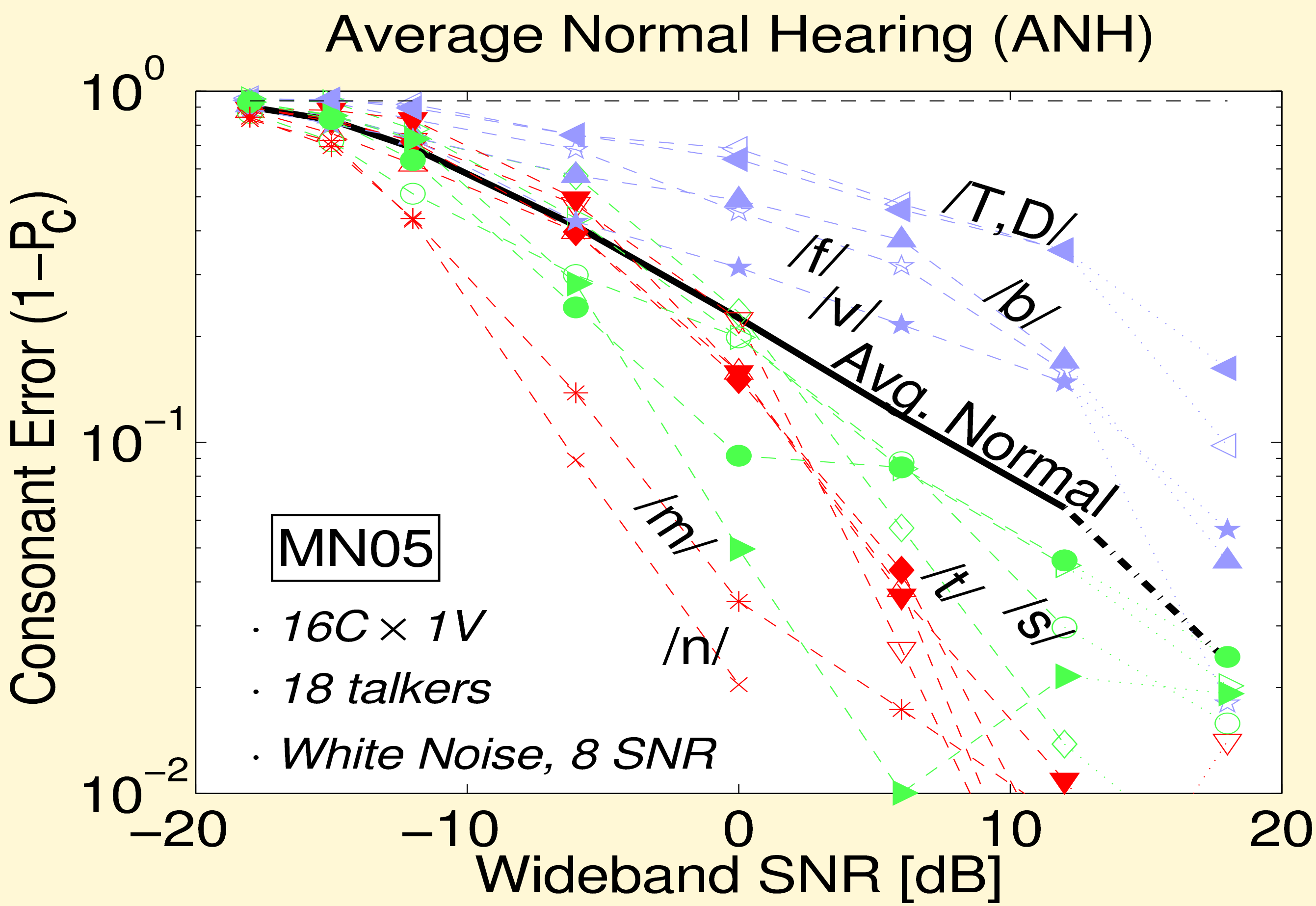
- The SIN$_c$ of averaging across consonants

- The SIN_c of averaging across consonants
  - The Consonant error $P_e(\mathit{SNR}|s)$ is heterogeneous!

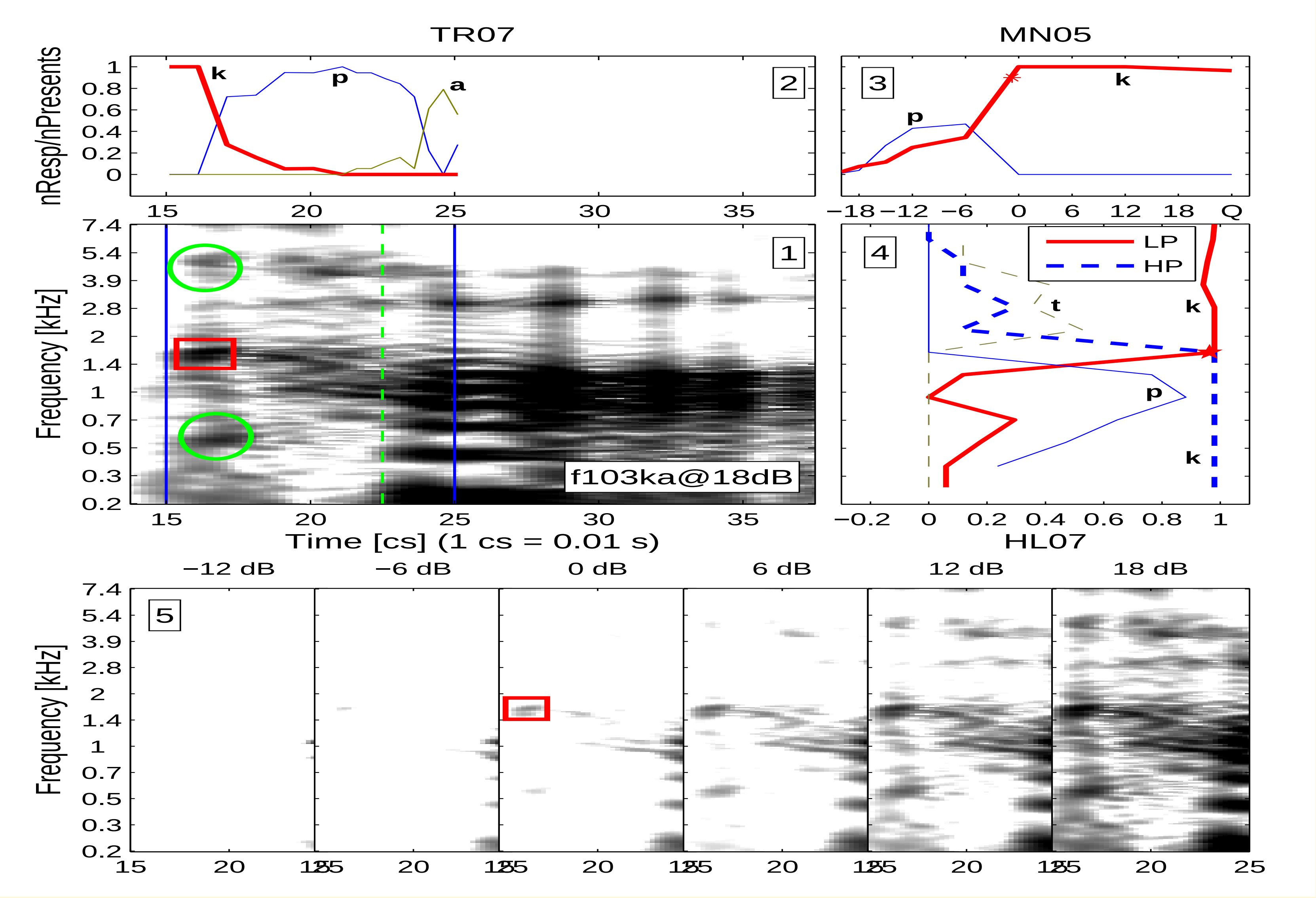
- The SIN_c of averaging across consonants
  - The Consonant error $P_e(SNR|s)$ is heterogeneous!



- The SIN_c of averaging across consonants
  - The Consonant error $P_e(SNR|s)$ is heterogeneous!

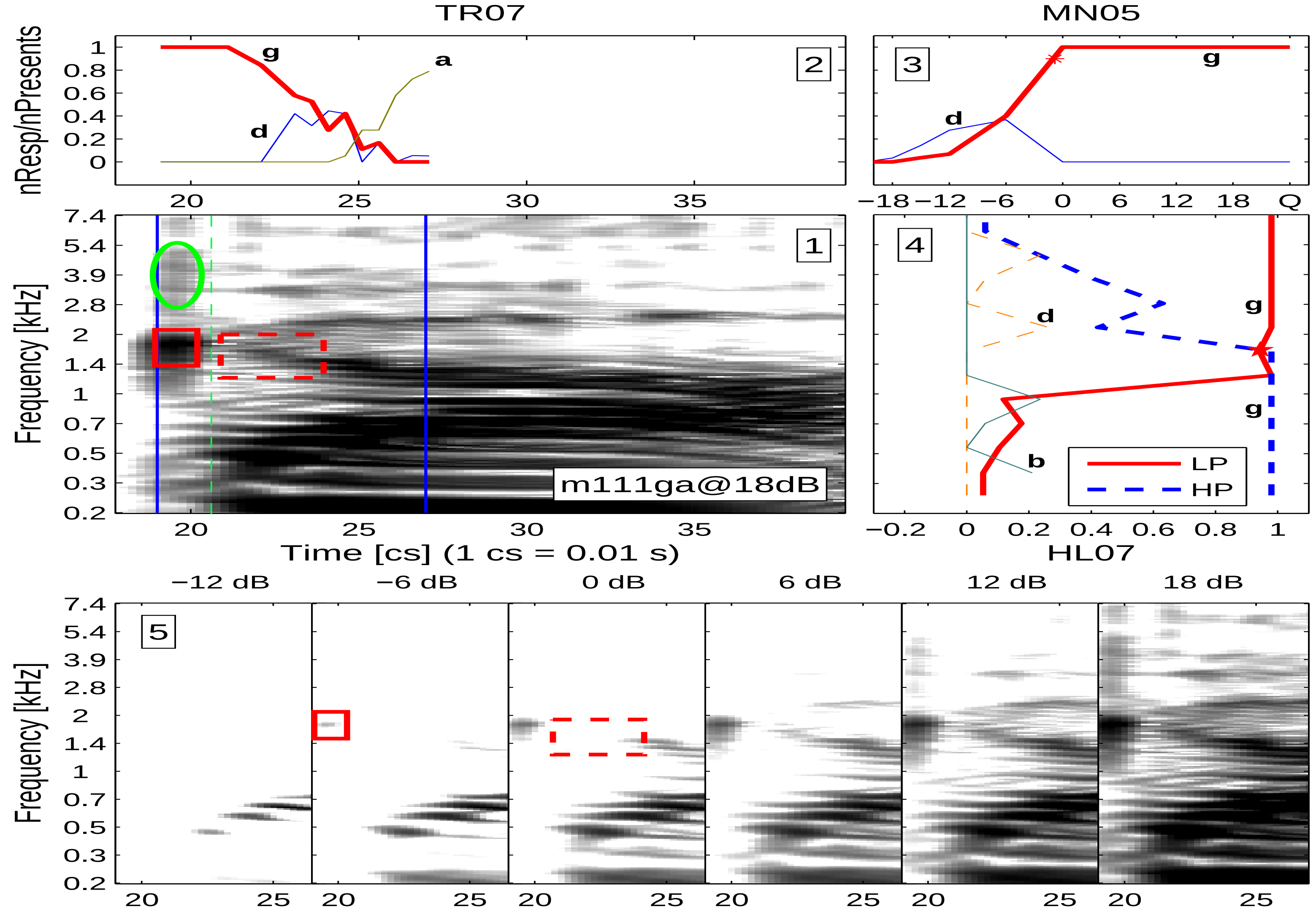


- Averaging obscures the multimodal consonant error distribution

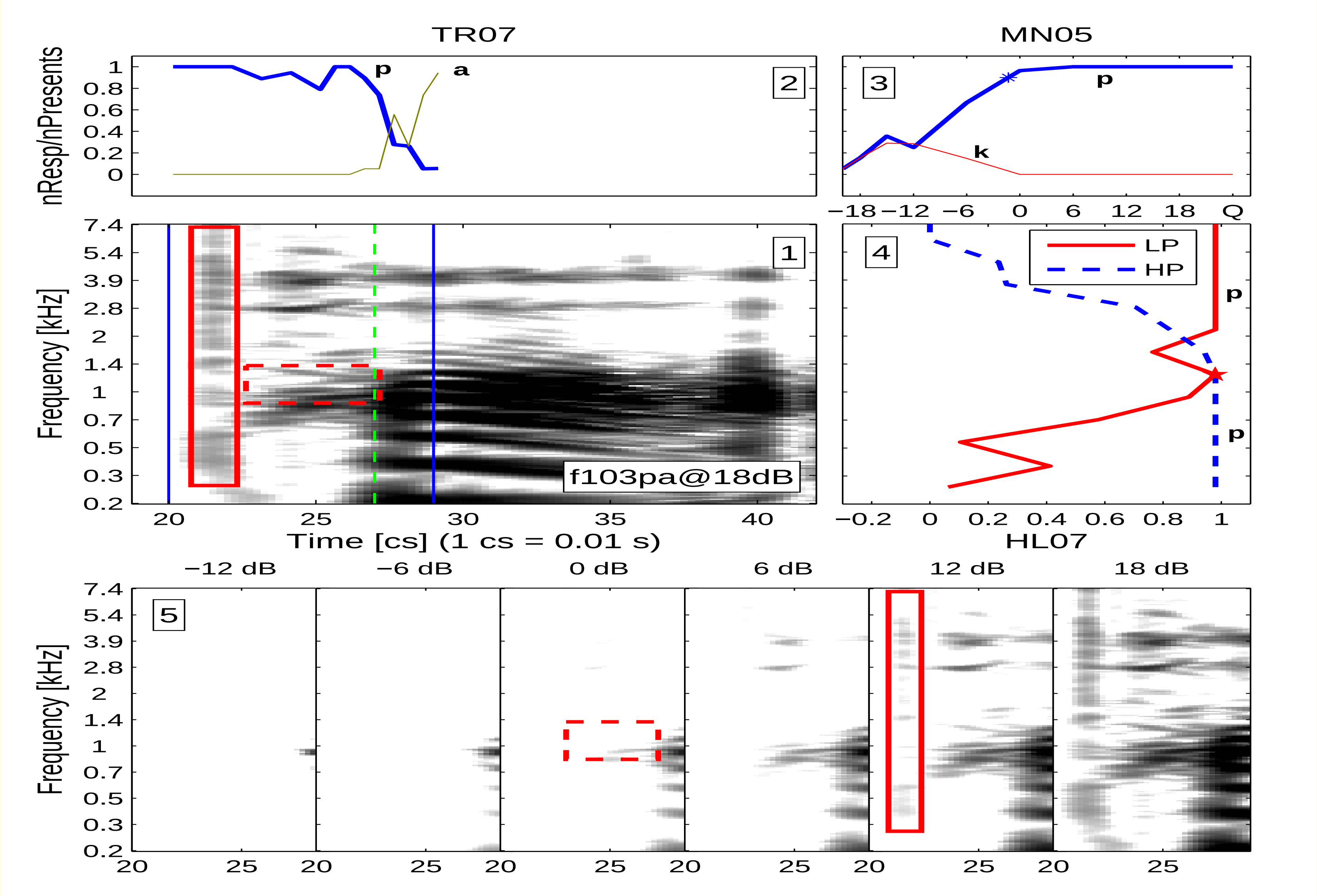


TR07
MN05
nResp/nPresents
k
p
a
2
3
1
4
5
Frequency [kHz]
f103ka@18dB
Time [cs] (1 cs = 0.01 s)
HL07
LP
HP
t
Q
−12 dB
−6 dB
0 dB
6 dB
12 dB
18 dB

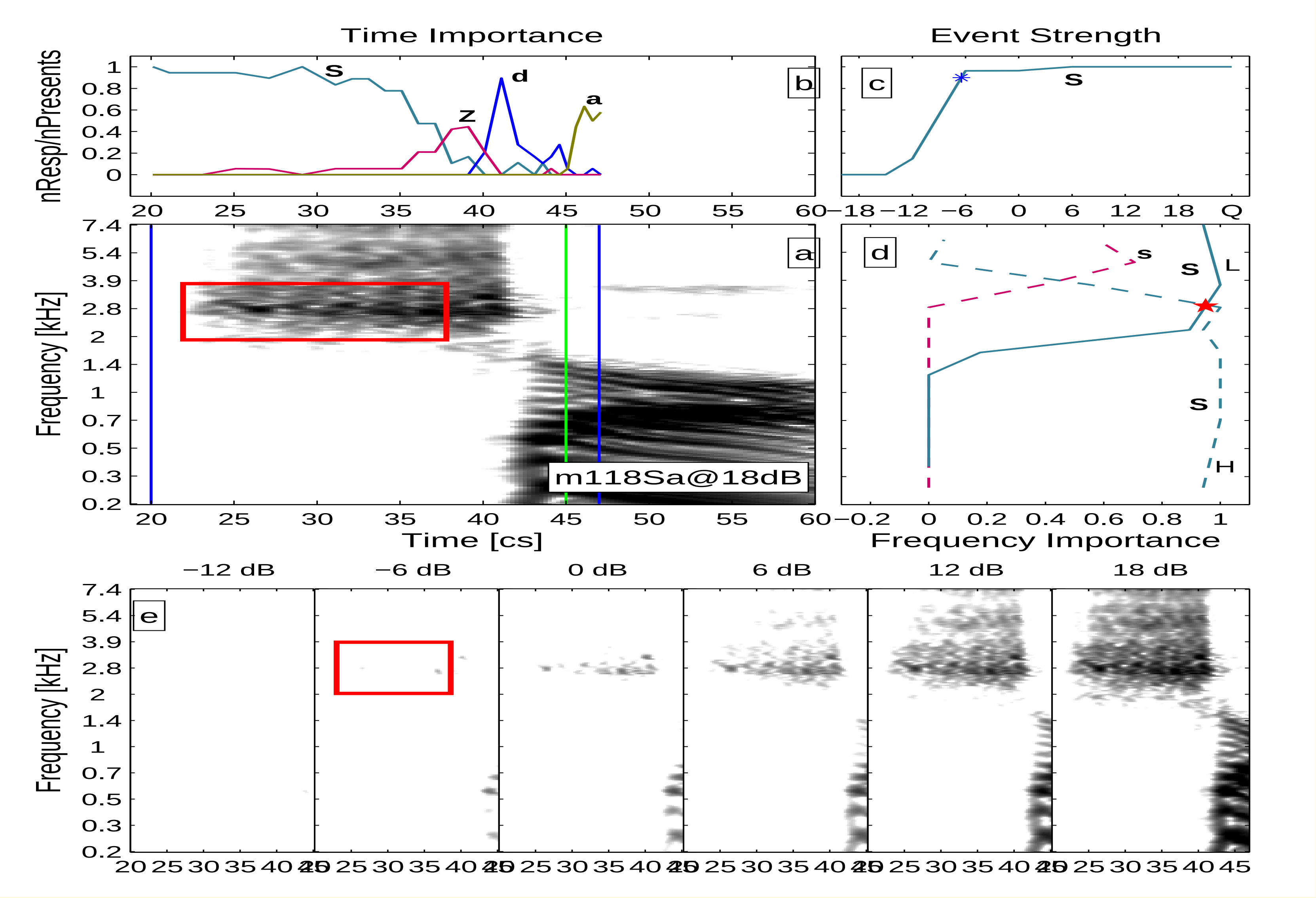
3DDS Method /ga/



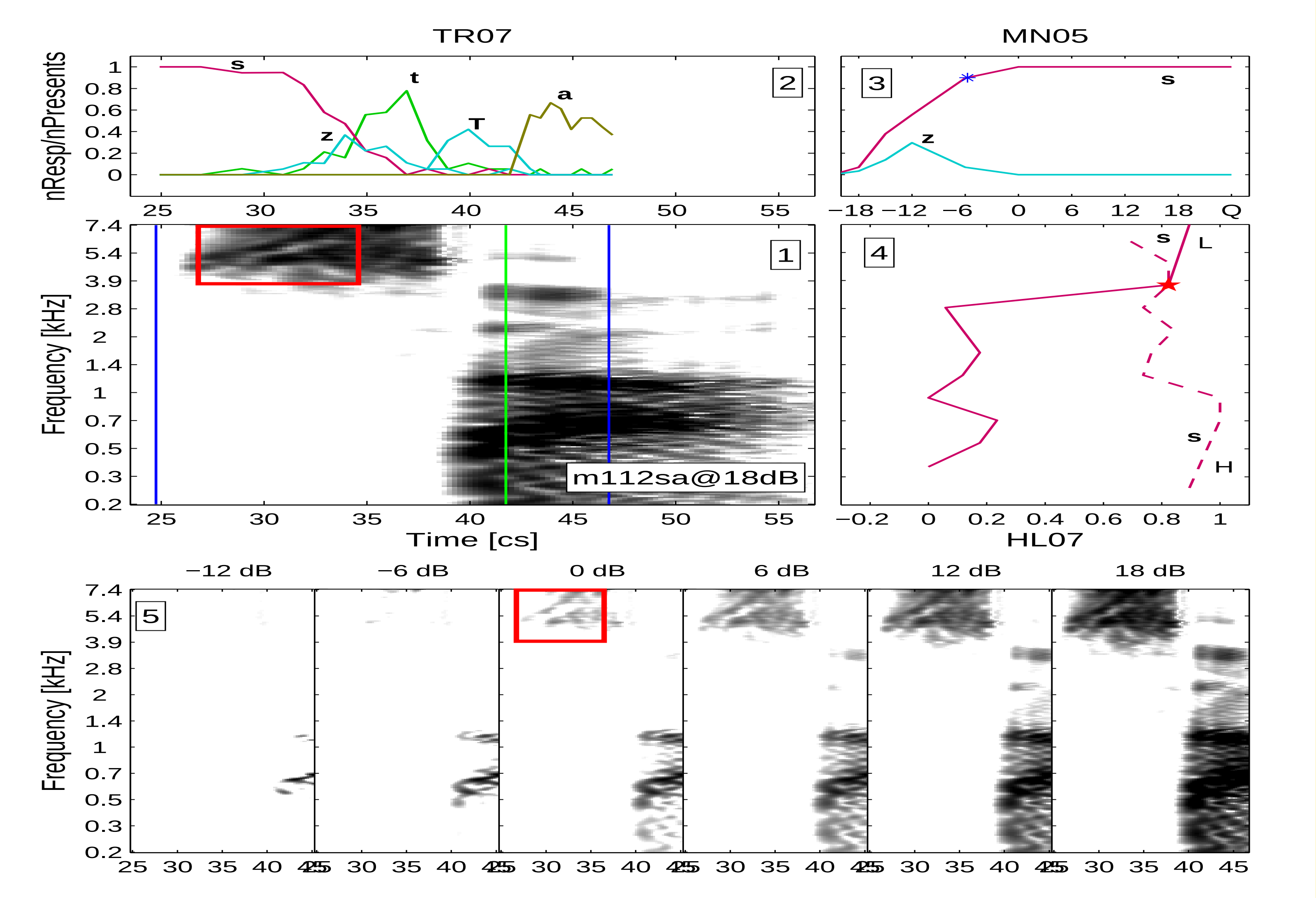
3DDS Method /pa/

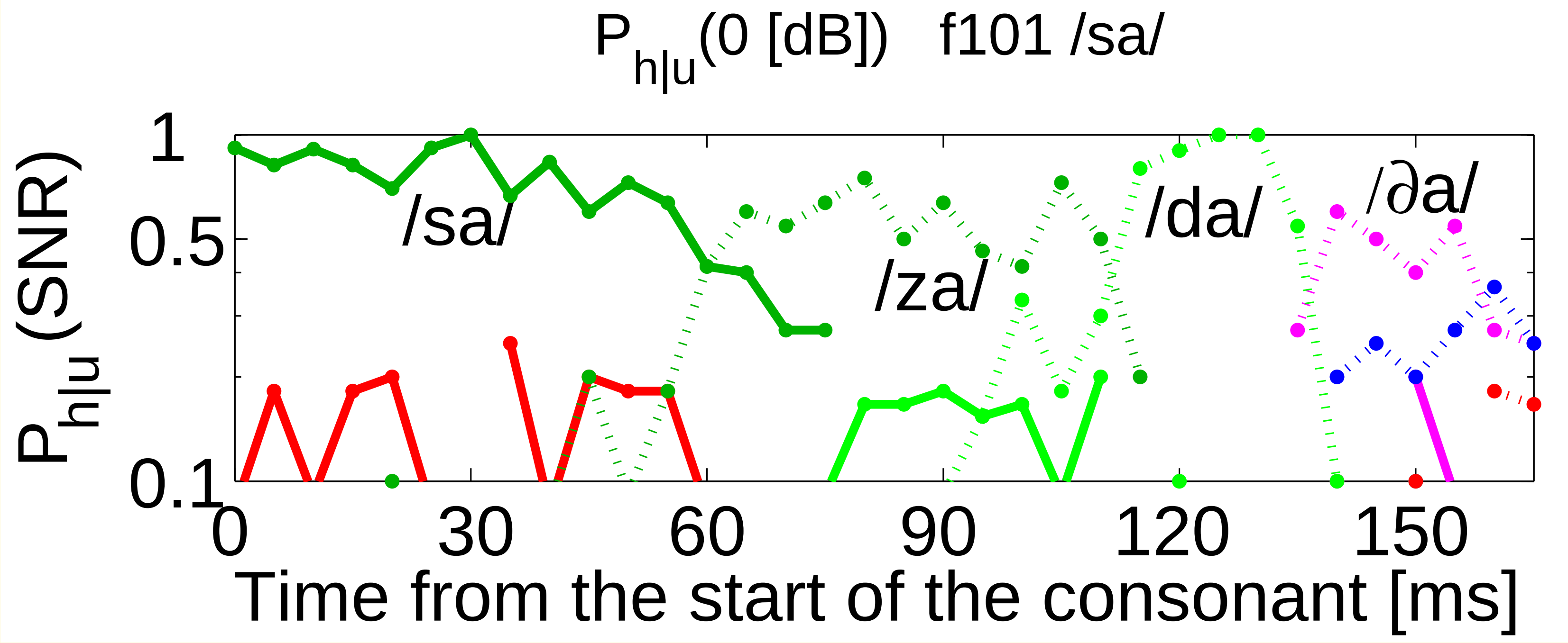


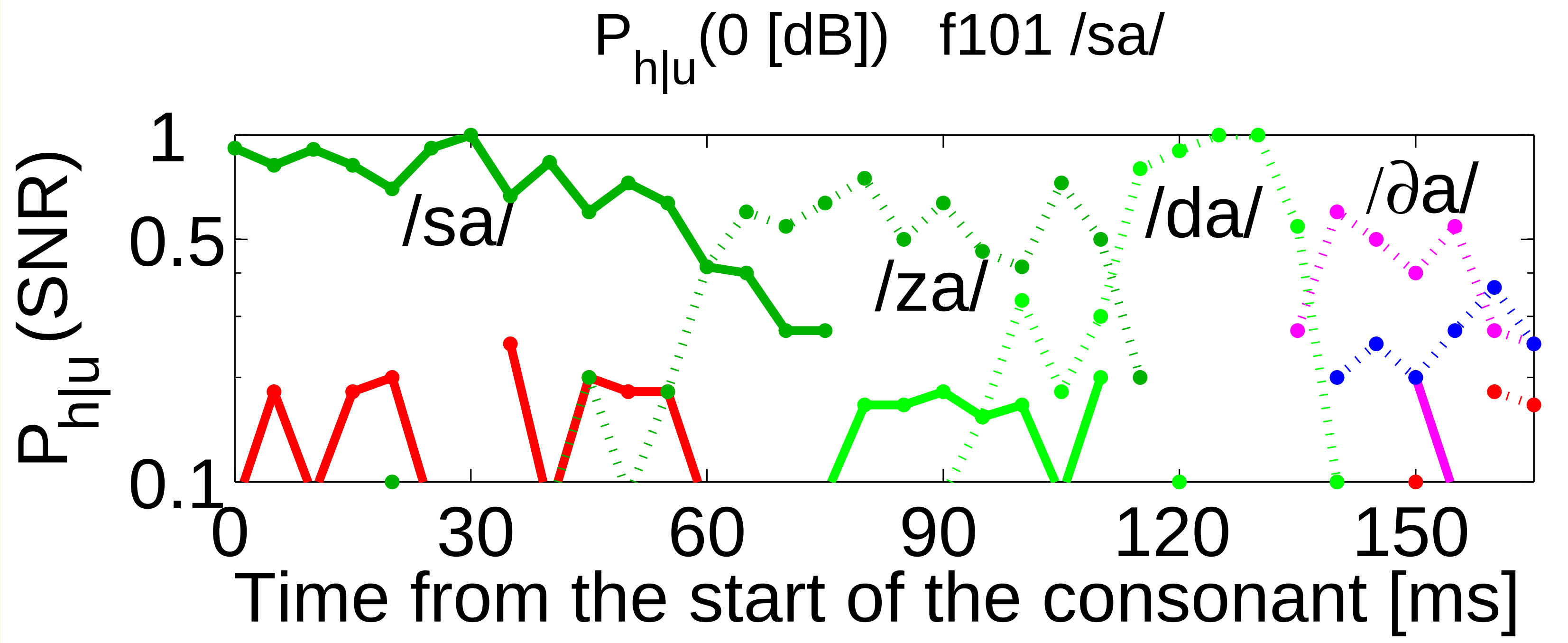
3^d-DS Method /ʃa/



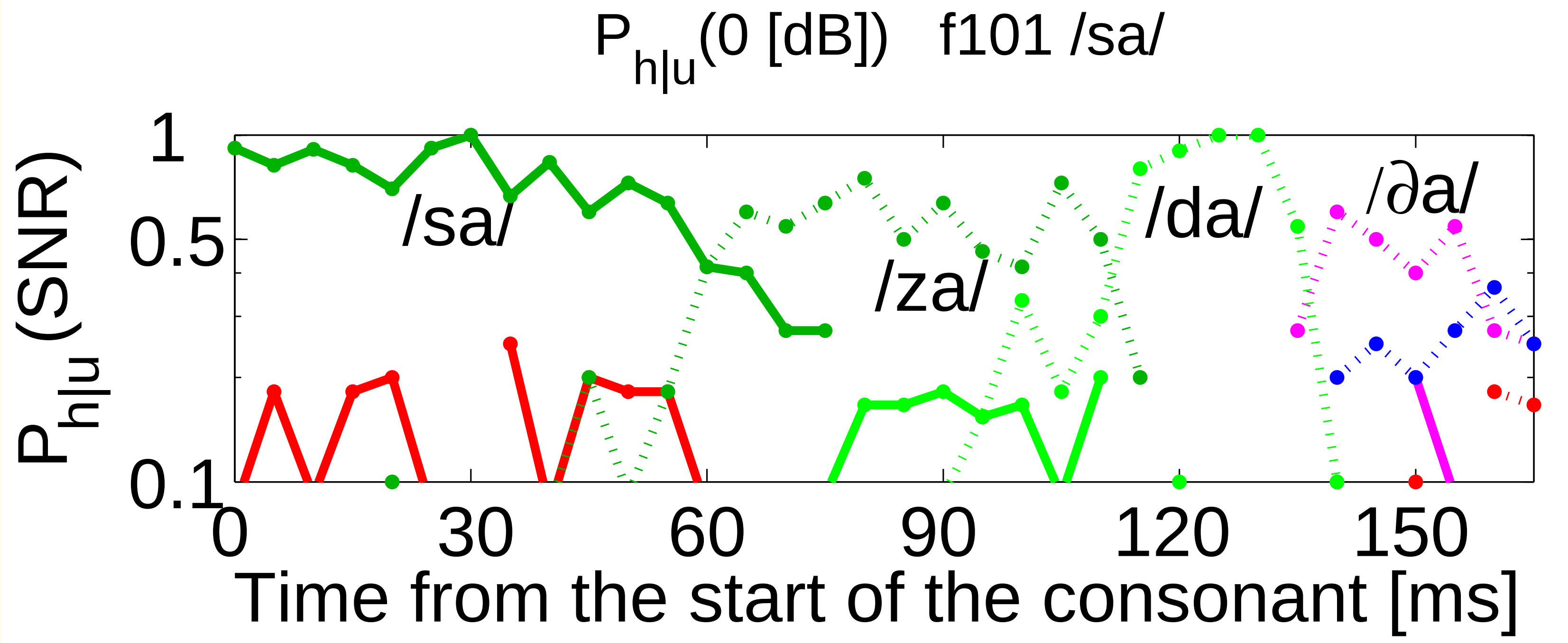
3DDS Method /sa/



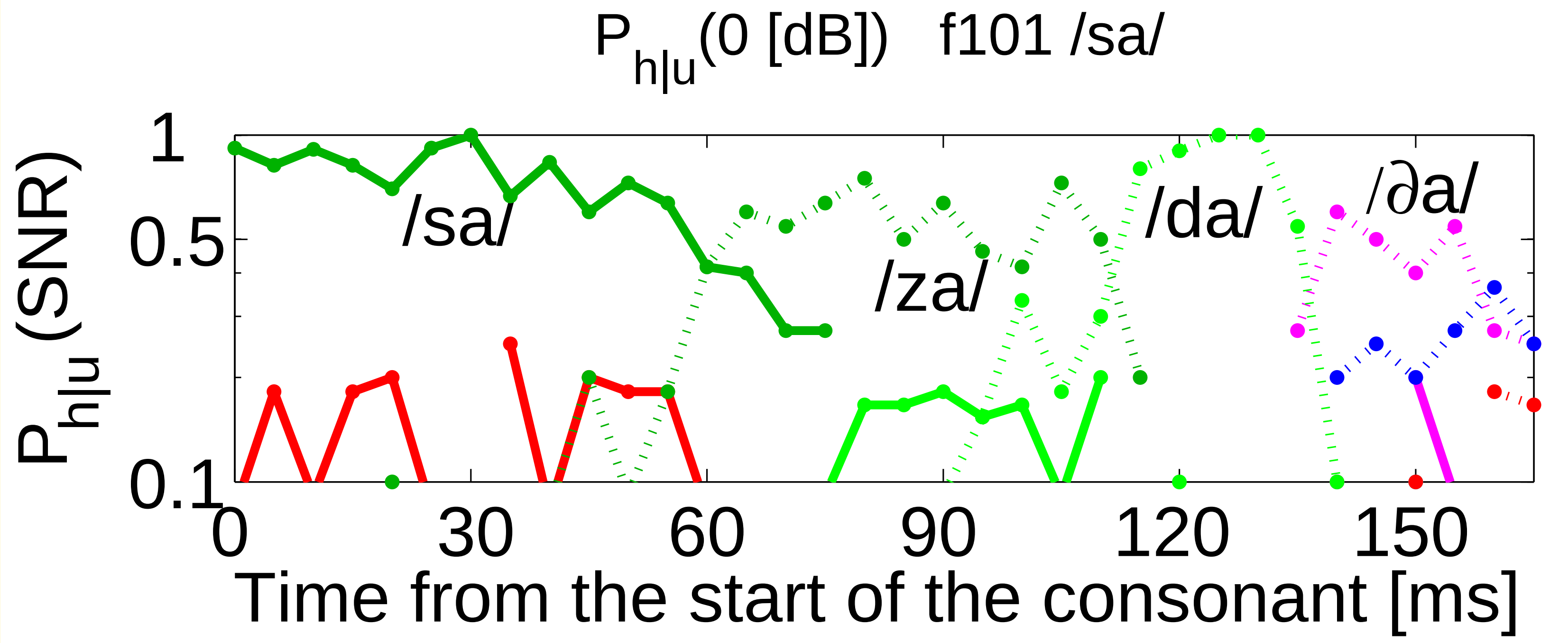

$P_{h|u}$(0 [dB]) f101 /sa/
$P_{h|u}$(SNR)
1
0.5
0.1
0
30
60
90
120
150
Time from the start of the consonant [ms]
/sa/
/za/
/da/
/∂a/



- NH responses to onset truncation /sɑ/



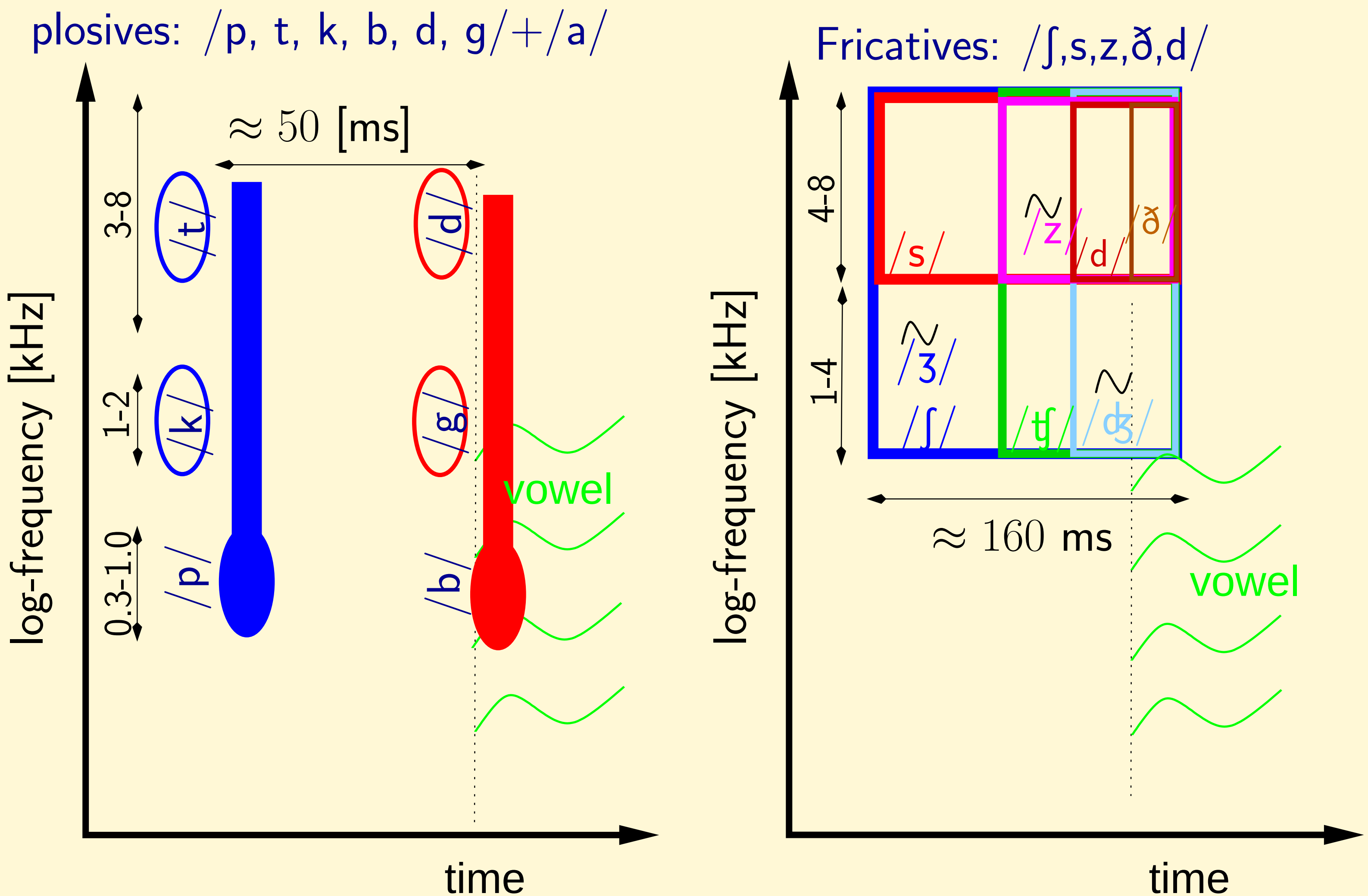
- NH responses to onset truncation /sɑ/
- Morphing from /sɑ/ → /zɑ/ → /dɑ/ → /ðɑ/



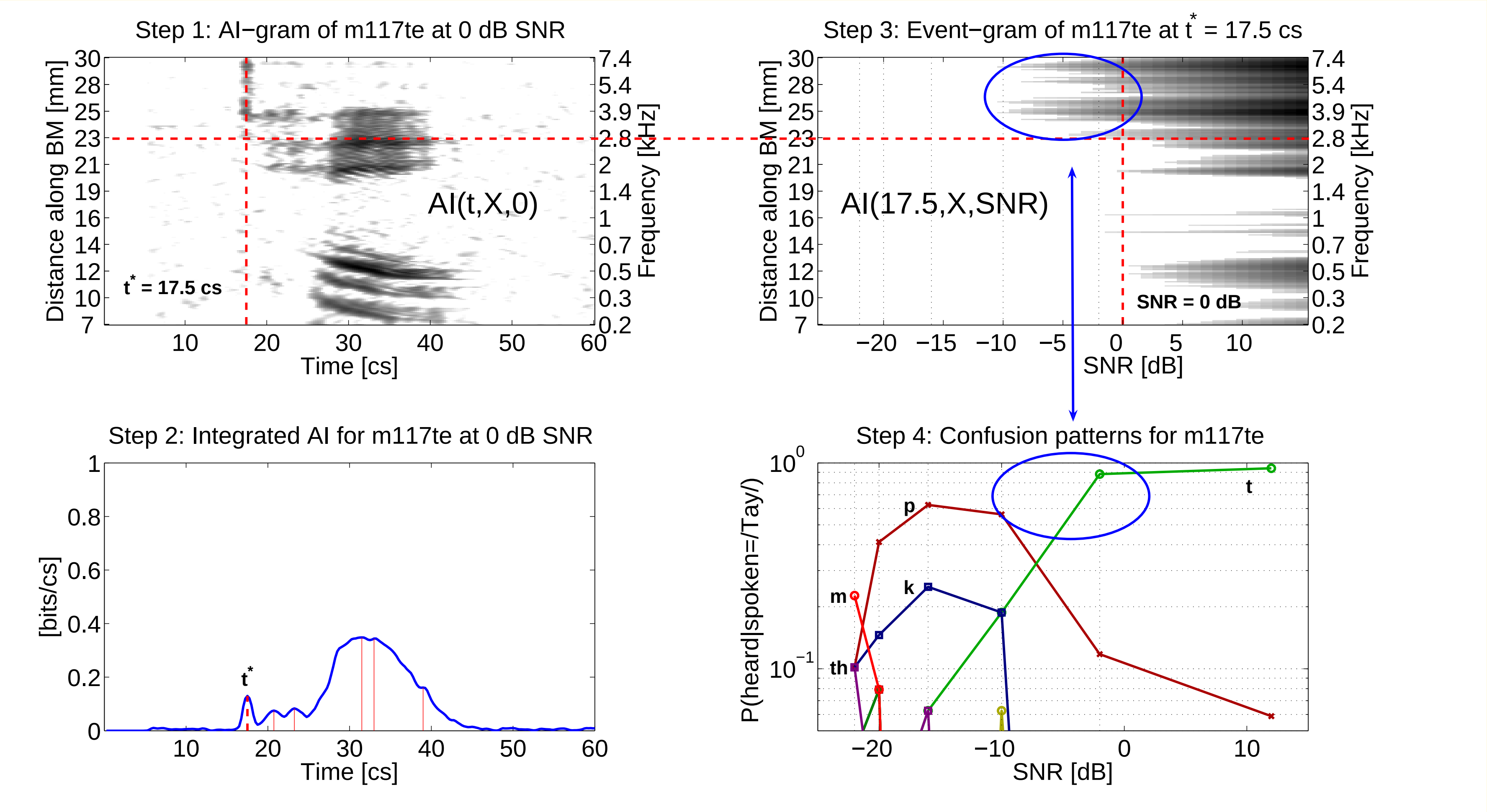
- NH responses to onset truncation /sɑ/
- Morphing from /sɑ/ → /zɑ/ → /dɑ/ → /ðɑ/
- Duration, low-frequency edge and F_0 modulations define a fricative cue

Summary of Consonant structure

- Time-frequency structure of plosives and fricatives

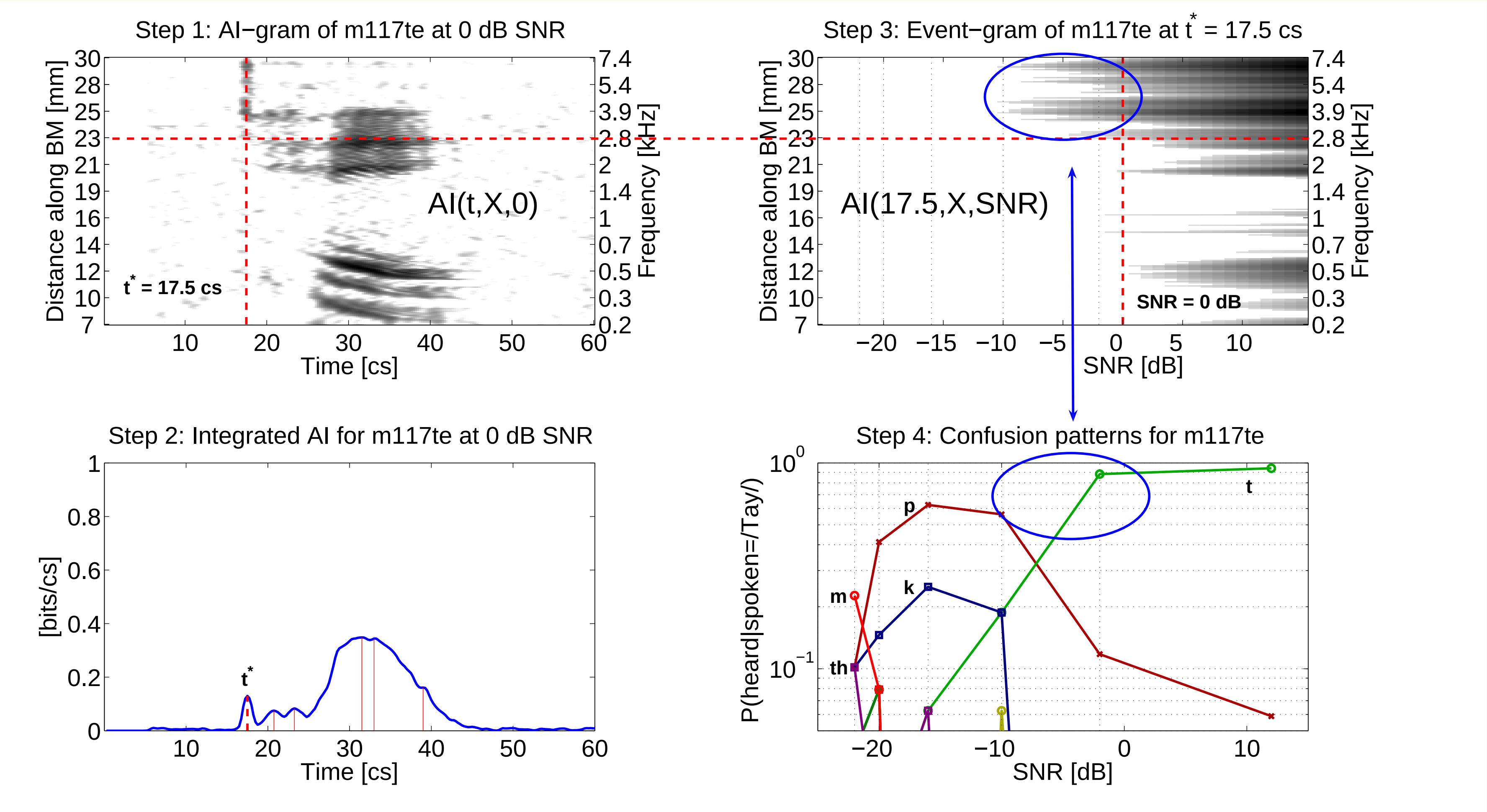


m117/te/ in speech-weighted noise



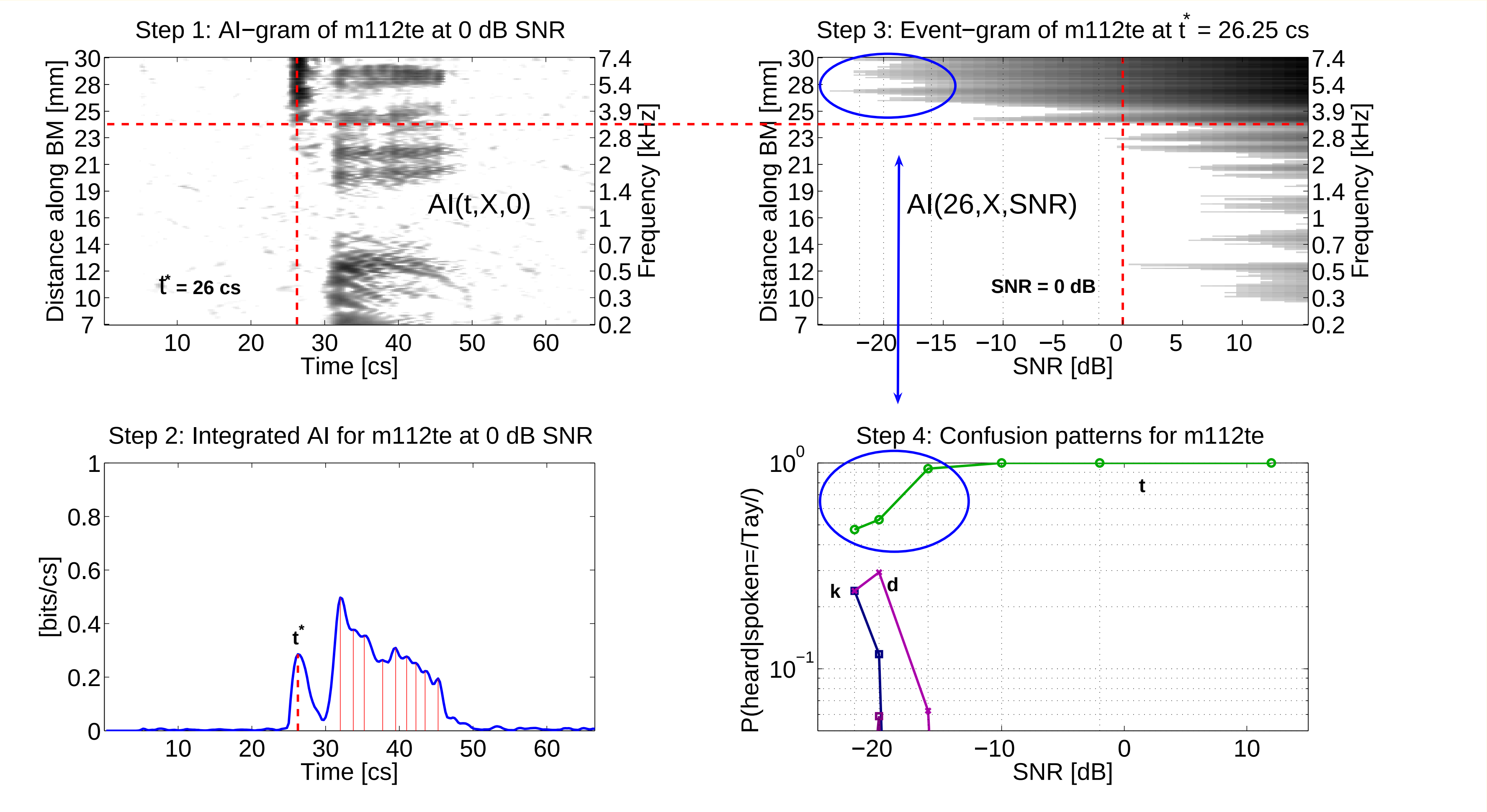
- /t/ confusion threshold at $P_c(SNR^*_{90} = -2)$ dB at 90% correct

m117/te/ in speech-weighted noise



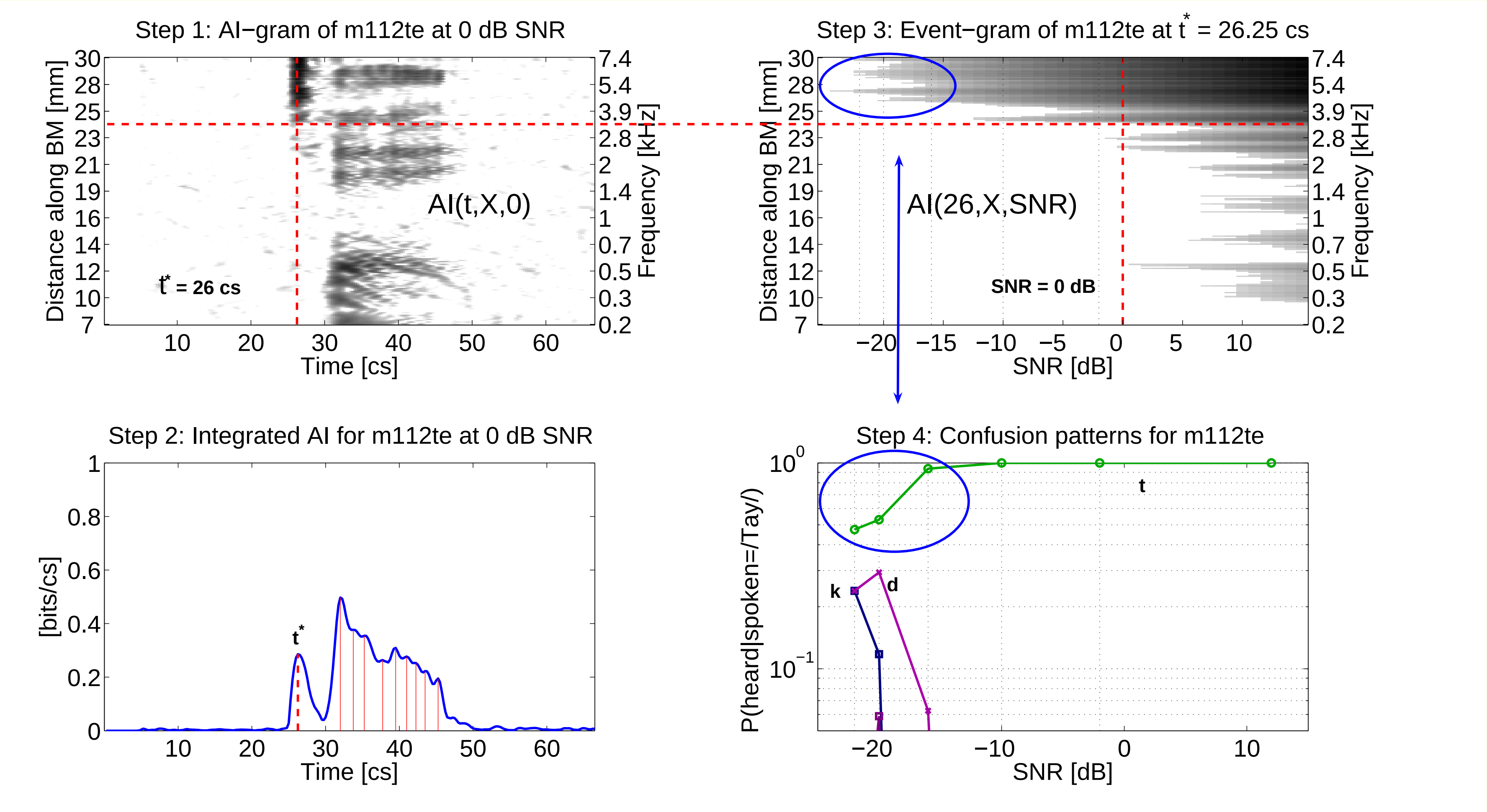
- /t/ confusion threshold at $P_c(SNR^*_{90} = -2)$ dB at 90% correct
- This is an example of a high-error /te/

m112/te/ in speech-weighted noise



- /t/ confusion threshold at $P_c(SNR^*_{90} = -16)$ dB at 90% correct

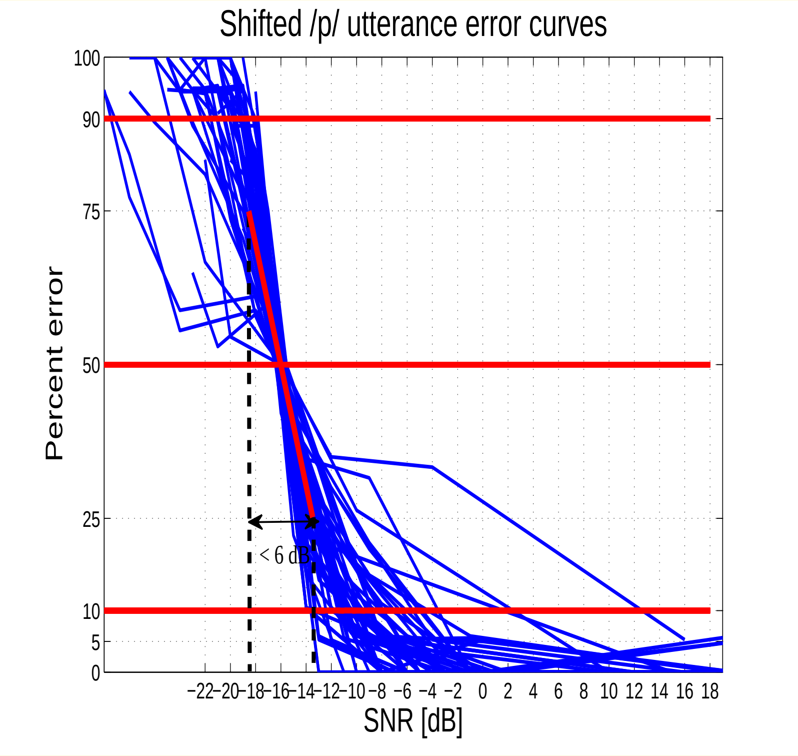
m112/te/ in speech-weighted noise



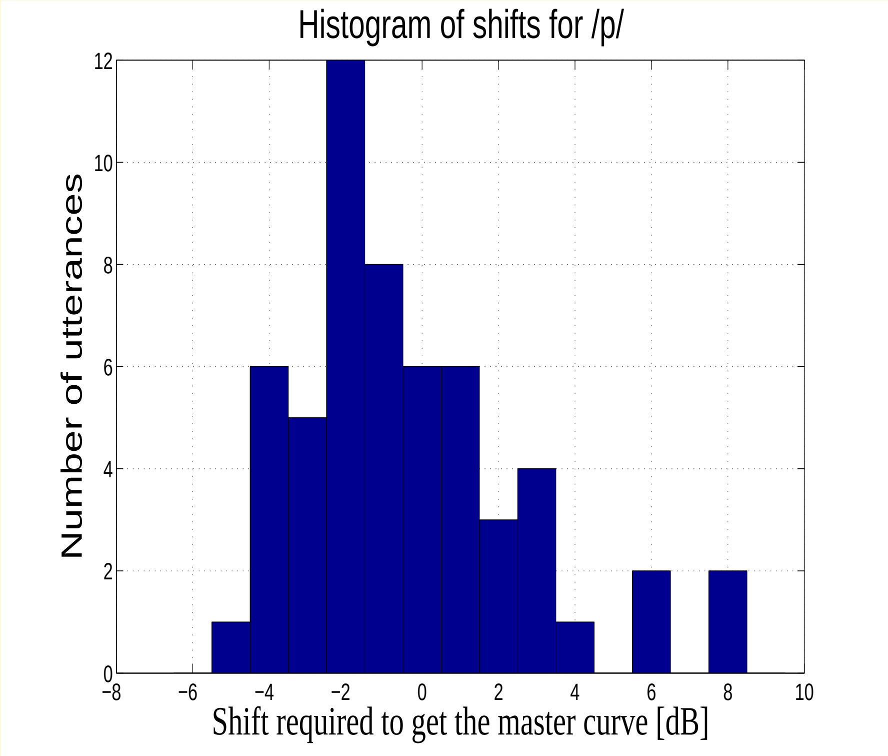
- /t/ confusion threshold at $P_c(SNR^*_{90} = -16)$ dB at 90% correct
- This is an example of a low-error /te/

Properties of $P_e(SNR)$ for /p/

- Error vs. SNR shifted to 50% threshold SNR^*_{50} (LEFT)
- Histogram of 50% error thresholds (RIGHT)
  - Sharp transition ⇒ Binary Plosive identification!



(a) $P_e(SNR - SNR^*_{50})$



(b) Distribution of SNR^*_{50}

- The average score is quantified by the articulation index

- The average score is quantified by the articulation index
  - The AI only describe the *average score*
  - The AI has a *huge* variance ($\Rightarrow$ low utility)

Conclusion NH ears

- The average score is quantified by the articulation index
  - The AI only describe the *average score*
  - The AI has a *huge* variance ($\Rightarrow$ low utility)
- The AI's variance depends on at least 2 factors:
  Singh Allen JASA, 2012
  - We have identified specific consonant cues

- The average score is quantified by the articulation index
  - The AI only describe the *average score*
  - The AI has a *huge* variance ($\Rightarrow$ low utility)
- The AI's variance depends on at least 2 factors:
  Singh Allen JASA, 2012
  - We have identified specific consonant cues
    1. Across consonant variance ($>$20 dB spread)
    2. Across talker dependent thresholds (SNR_{50} 20 dB)

- The average score is quantified by the articulation index
  - The AI only describe the *average score*
  - The AI has a *huge* variance ($\Rightarrow$ low utility)
- The AI's variance depends on at least 2 factors:
  Singh Allen JASA, 2012
  - We have identified specific consonant cues
    1. Across consonant variance ($>$20 dB spread)
    2. Across talker dependent thresholds (SNR_{50} 20 dB)
- I have not yet discussed the HI token dependence

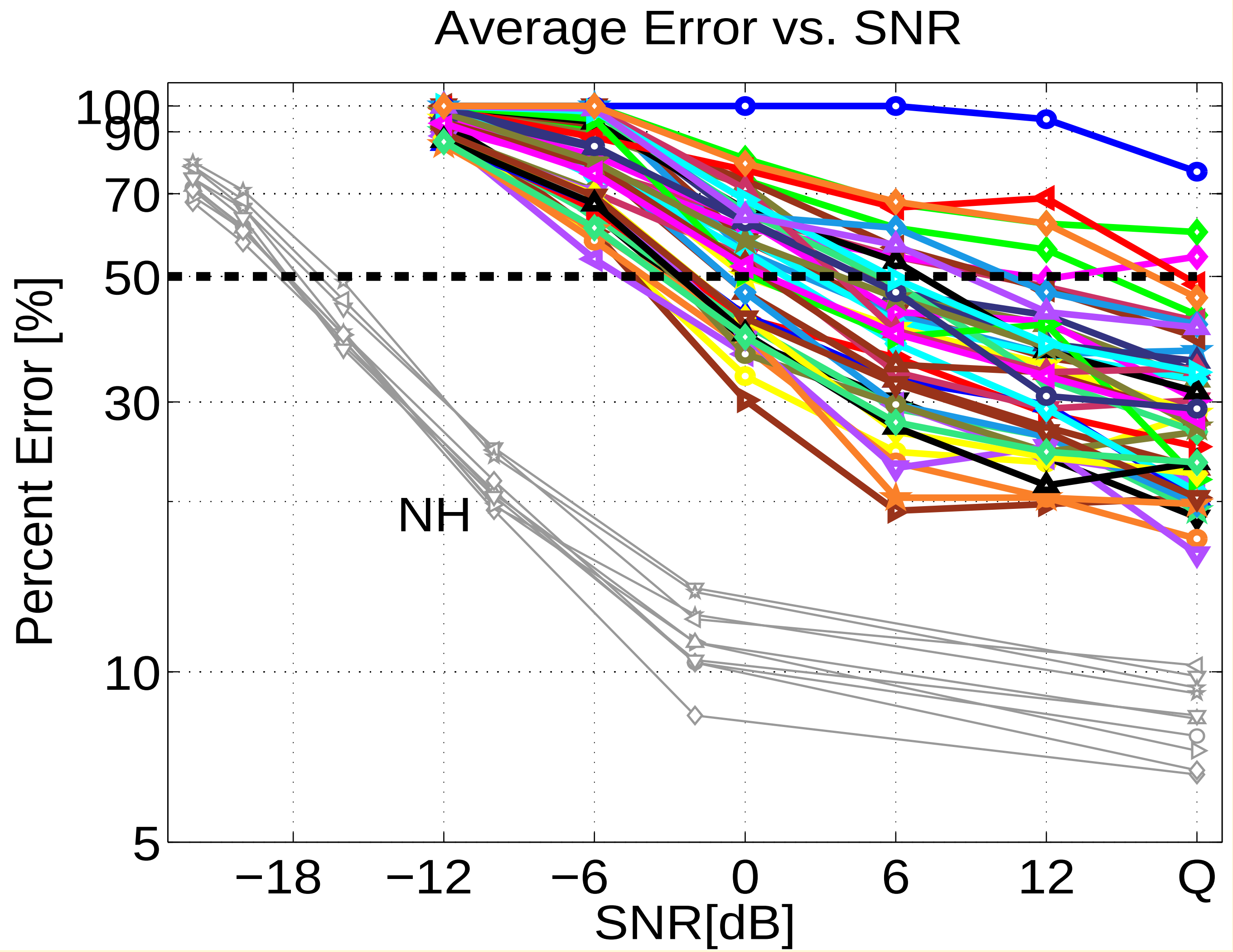
Outline

1. Intro + Objectives 2 mins
   - Research objectives
2. Historical overview 4 mins $\Sigma 6$
   - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
   - Speech-feature studies ($>$1950)
3. Methods 8 mins $\Sigma 14$
   - Channel capacity and the Articulation Index
   - Psychophysics of speech/AIgram/3DDS
4. Results with NH ears 10 mins $\Sigma 24$
   - Binary features; Confusions; Primes and Morphs;
5. Results with HI ears 20 mins $\Sigma 44$
   - Individual differences of consonant confusions
6. Summary + Conclusions 6 mins $\Sigma 50$

Yogi Berra Quote:

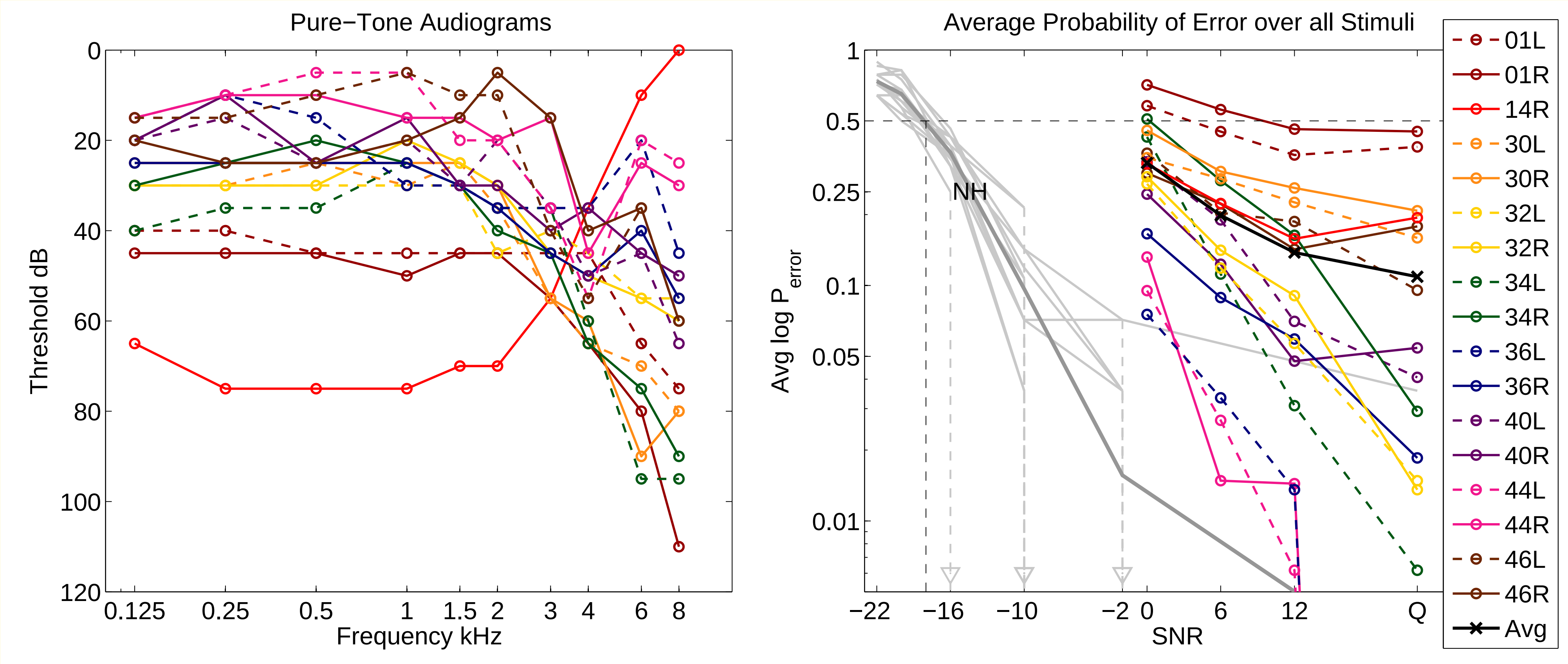
- "You can observe a lot by watching."

- Very large spread of error vs. SNR



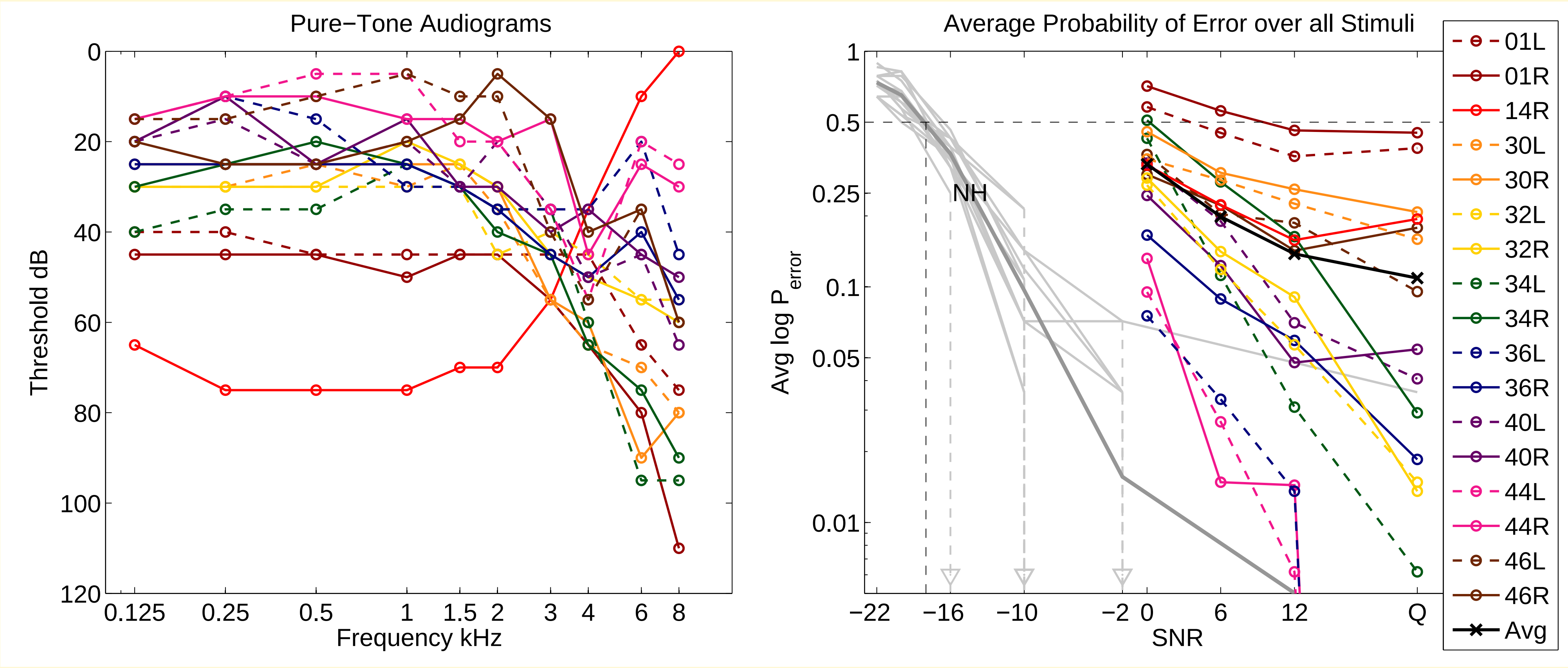
PTA & Average CV error over 17 HI ears

- 14 Miller-Nicely *zero-error* CV tokens:
  /p,t,k, b,d,g, f,s,ʃ, v,z,ʒ, m,n/+/a/



PTA & Average CV error over 17 HI ears

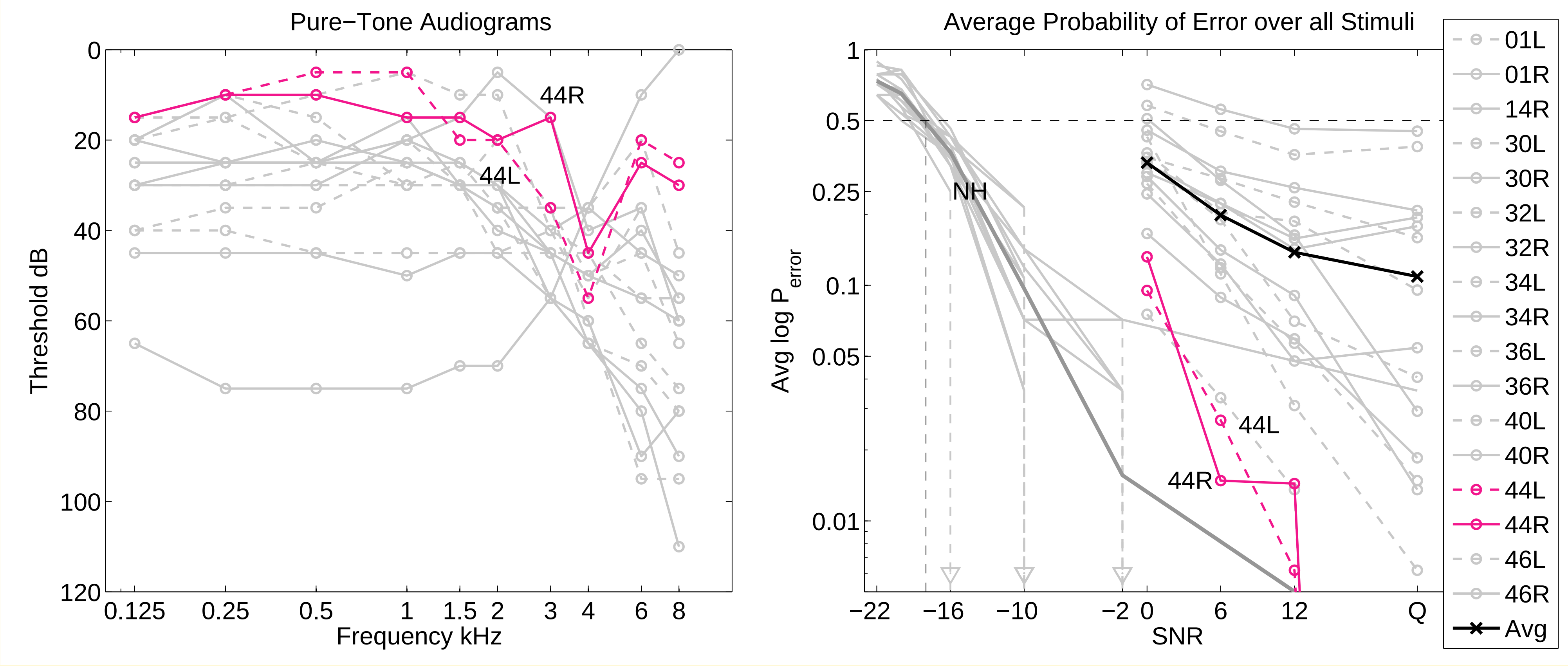
- 14 Miller-Nicely *zero-error* CV tokens:
  /p,t,k, b,d,g, f,s,ʃ, v,z,ʒ, m,n/+/a/



- 2 tokens for each CV (1-male & 1-female)

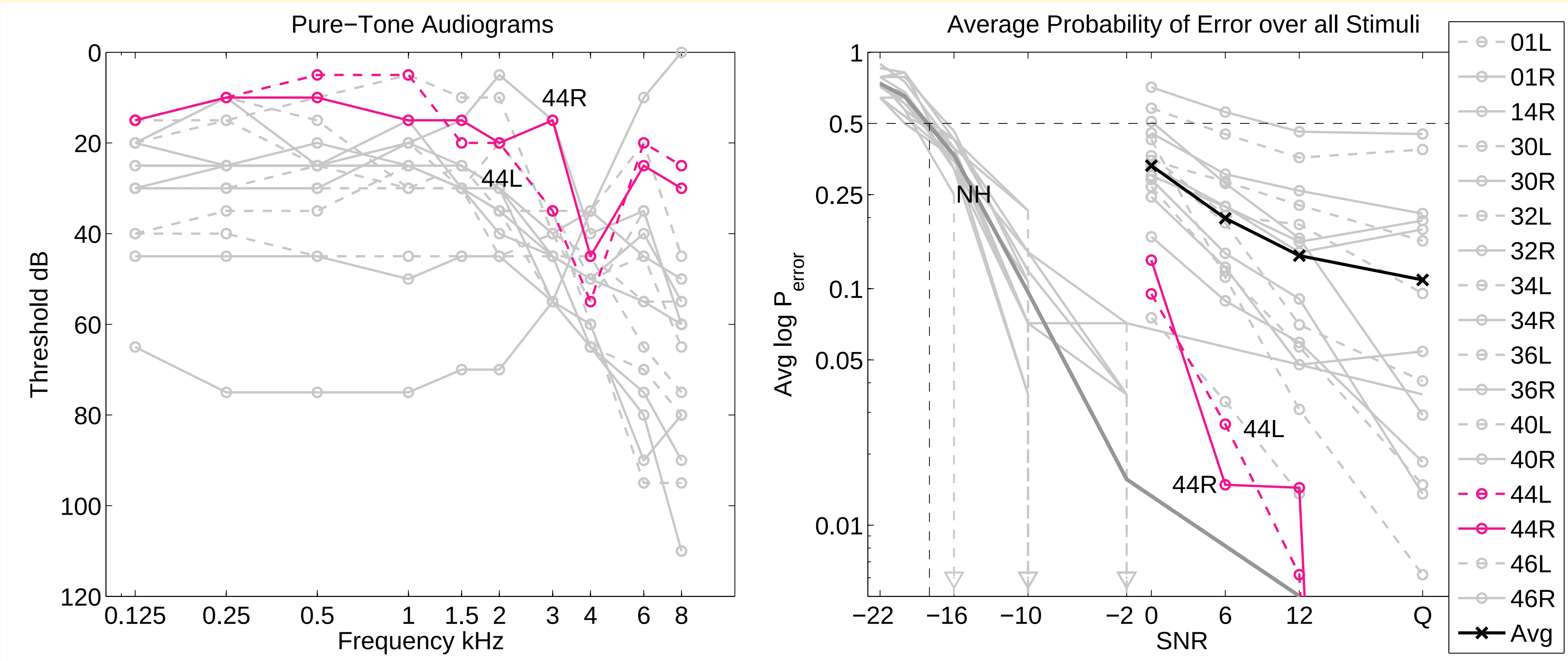
PTA & Average CV error – 44-L/R

- Subject 44-L/R (Left/Right) is our "best" listener!



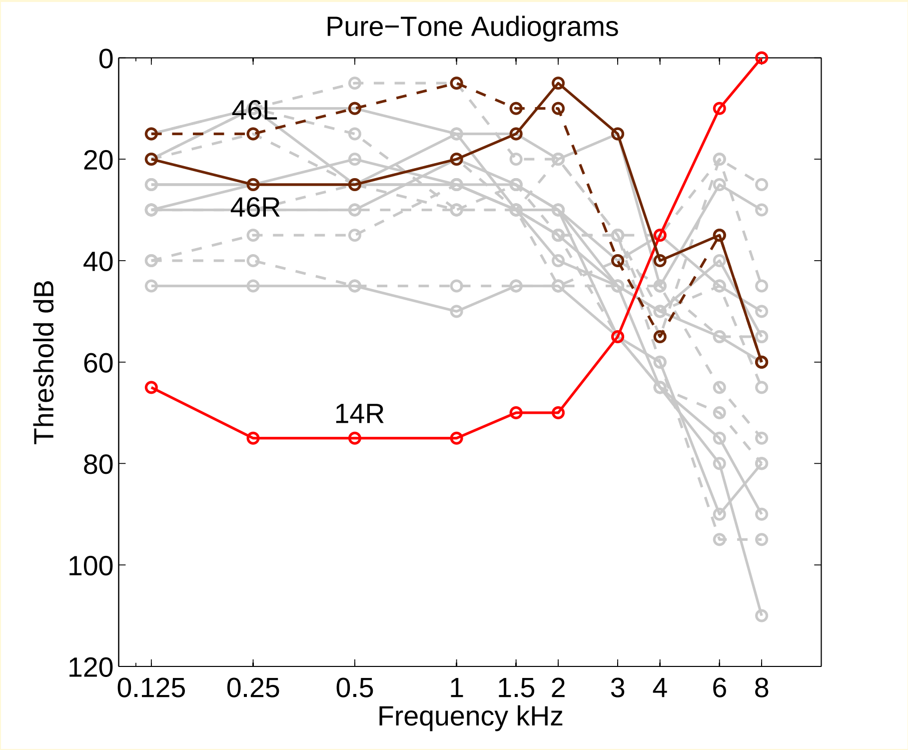
PTA & Average CV error – 44-L/R

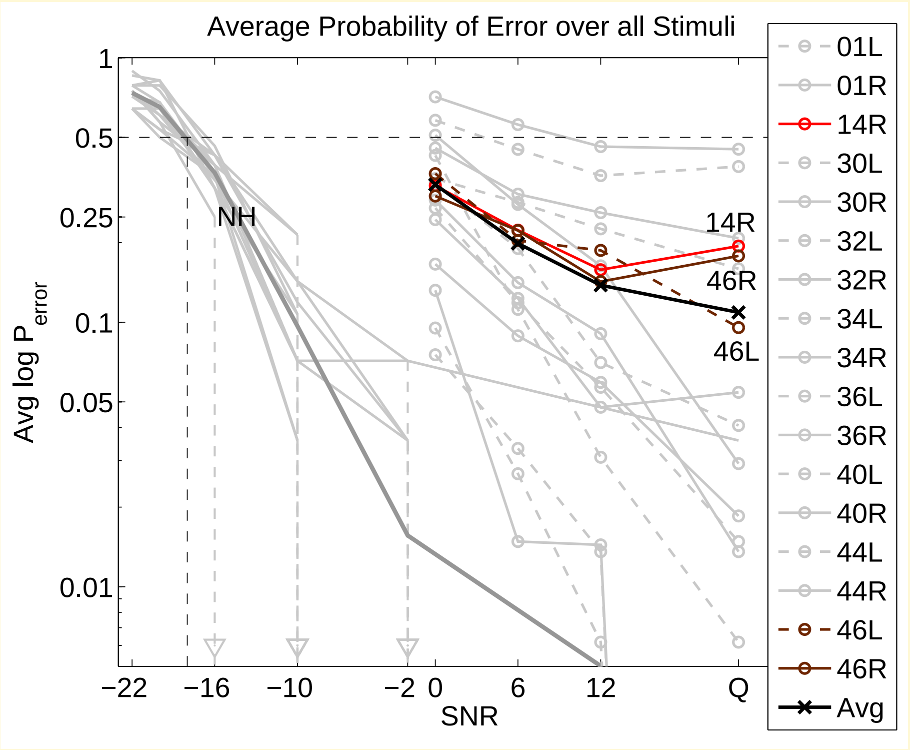
- Subject 44-L/R (Left/Right) is our "best" listener!



If only life were always this simple!

- Very different HL with nearly identical average speech scores

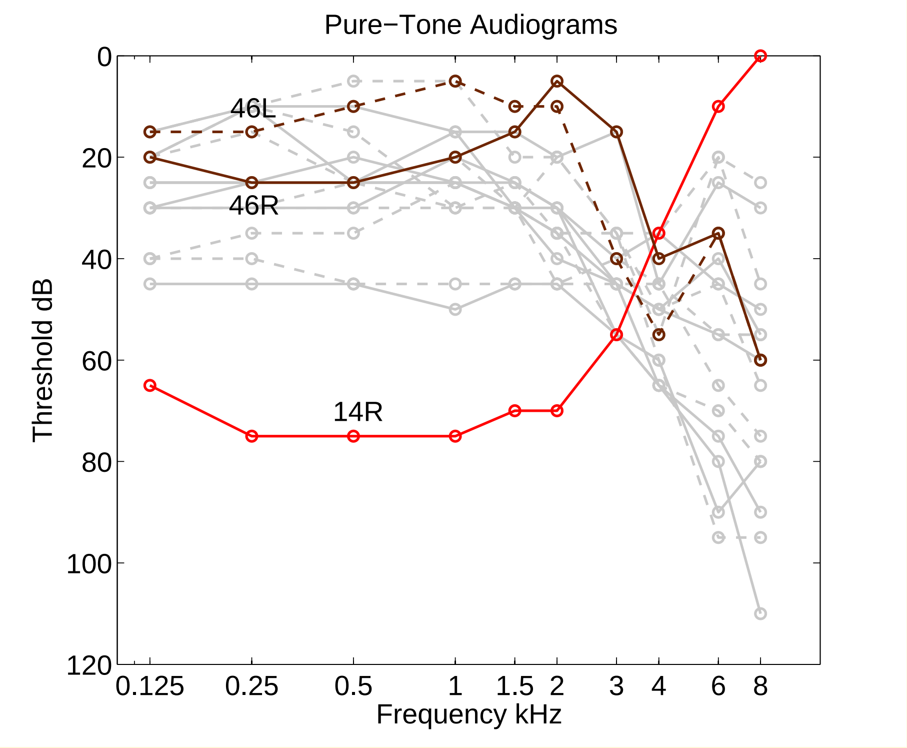


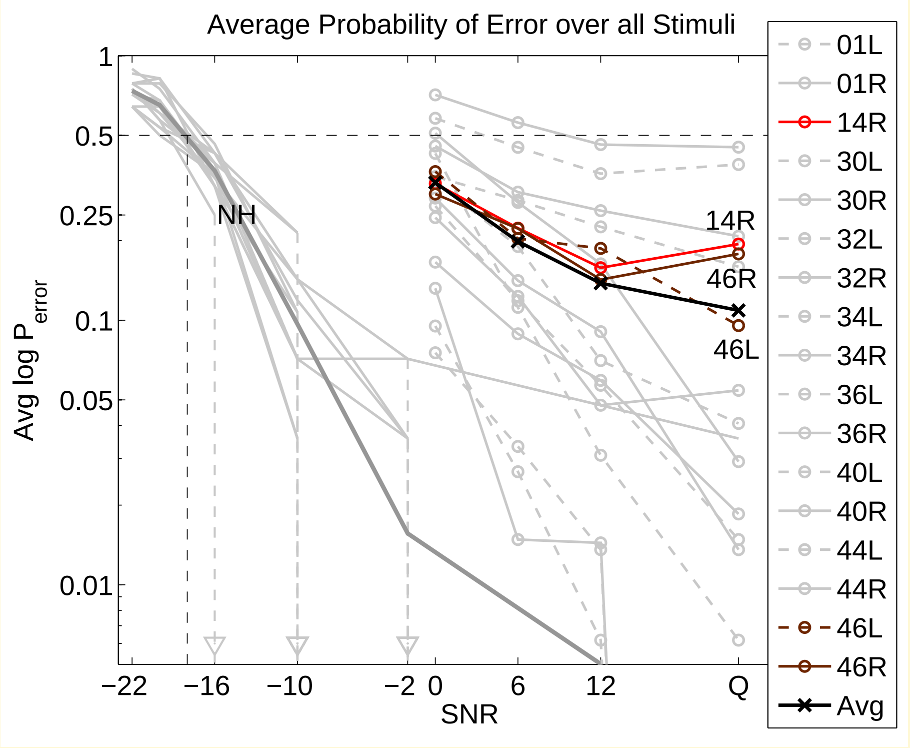


  - Average speech scores show no correlation with hearing-level,

PTA & Average CV error: 14R, 46L/R

- Very different HL with nearly identical average speech scores

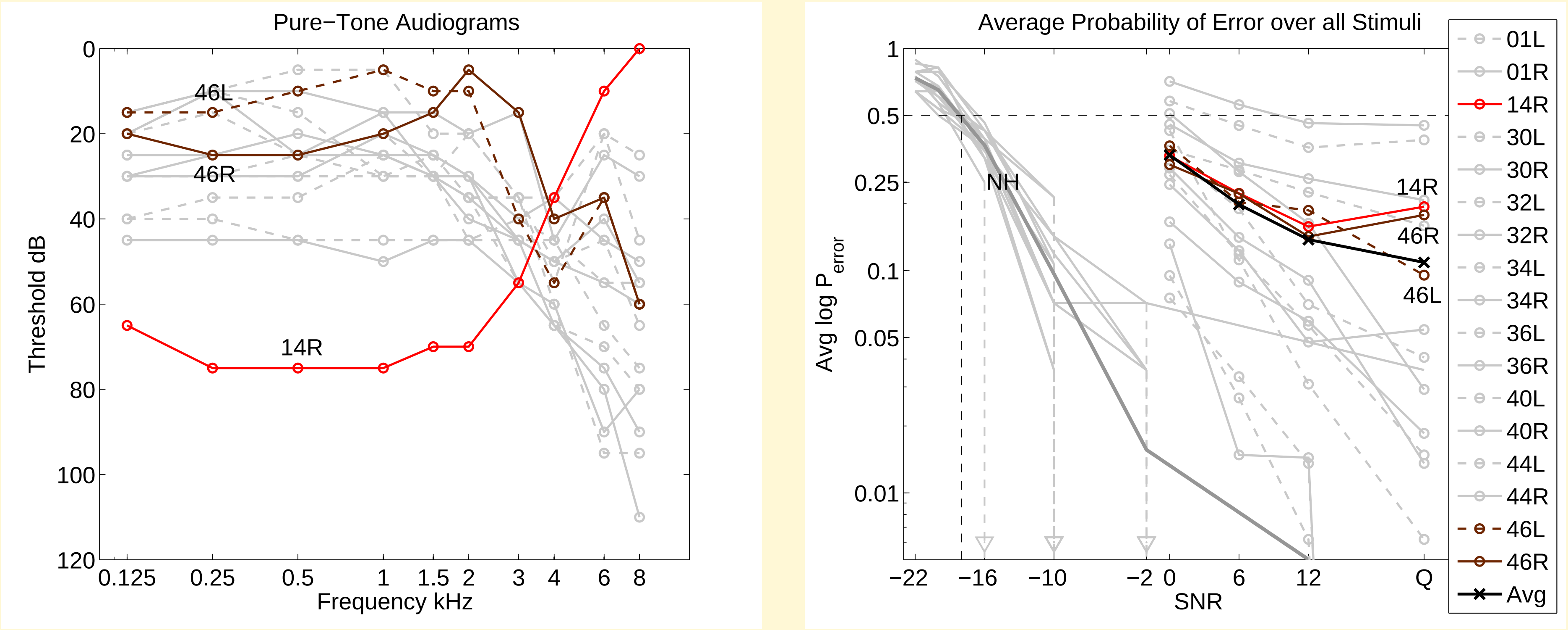




- Average speech scores show no correlation with hearing-level, with massively different audiograms!

PTA & Average CV error: 14R, 46L/R

- Very different HL with nearly identical average speech scores

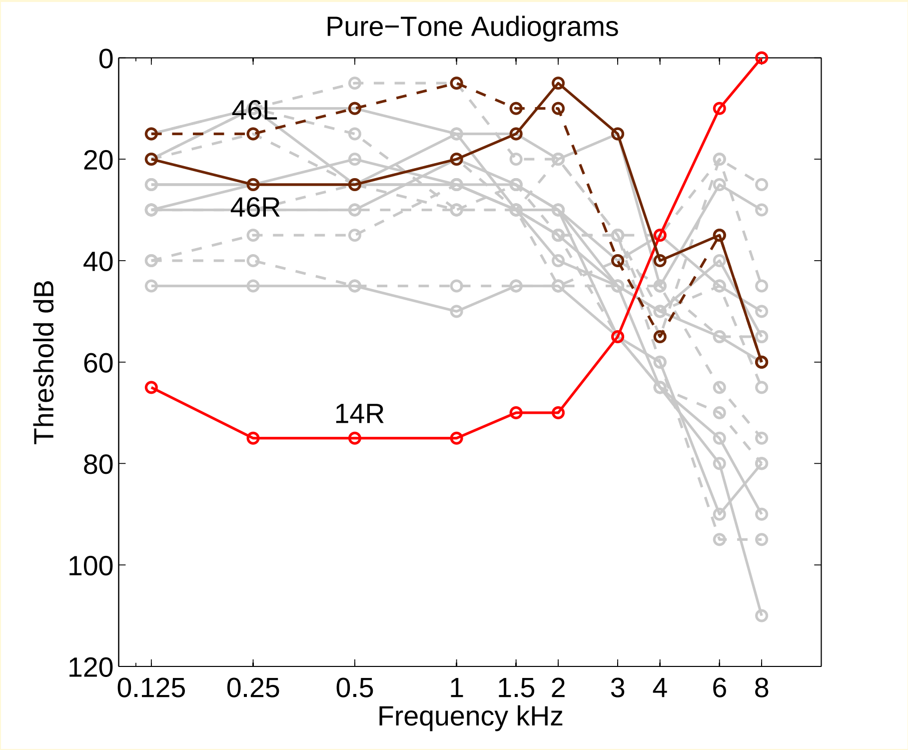


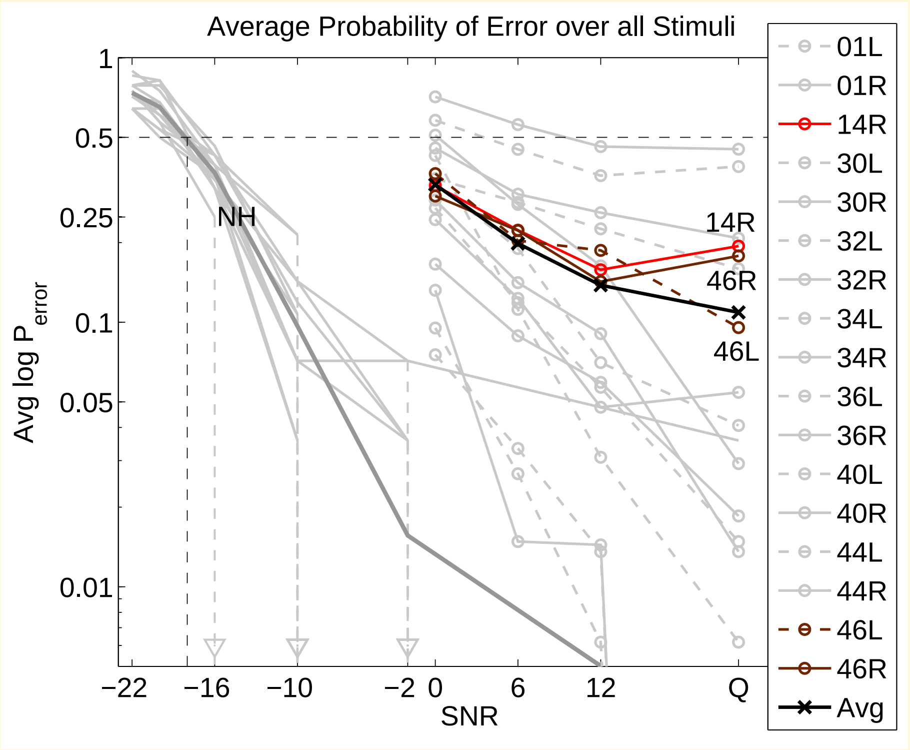
  - Average speech scores show no correlation with hearing-level, with massively different audiograms!

- Why? . . .

PTA & Average CV error: 14R, 46L/R

- Very different HL with nearly identical average speech scores



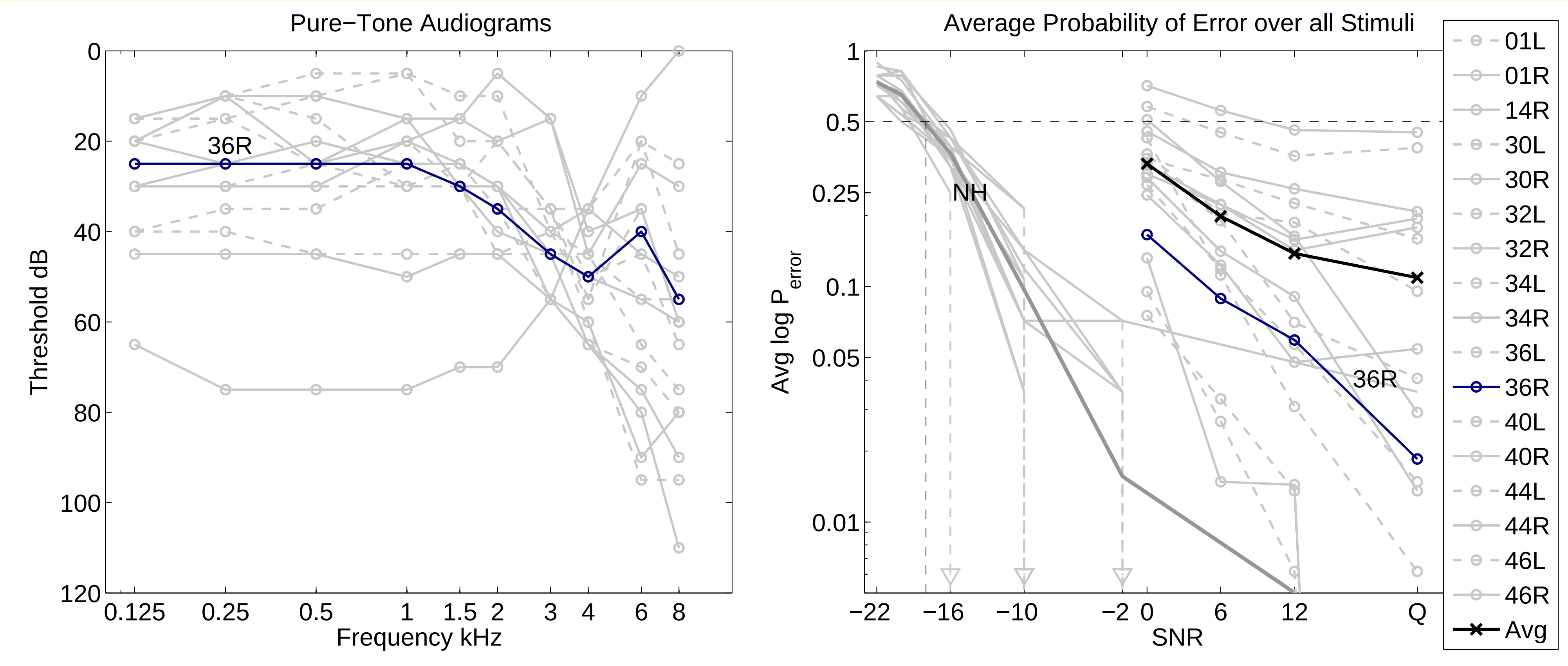


  - Average speech scores show no correlation with hearing-level, with massively different audiograms!

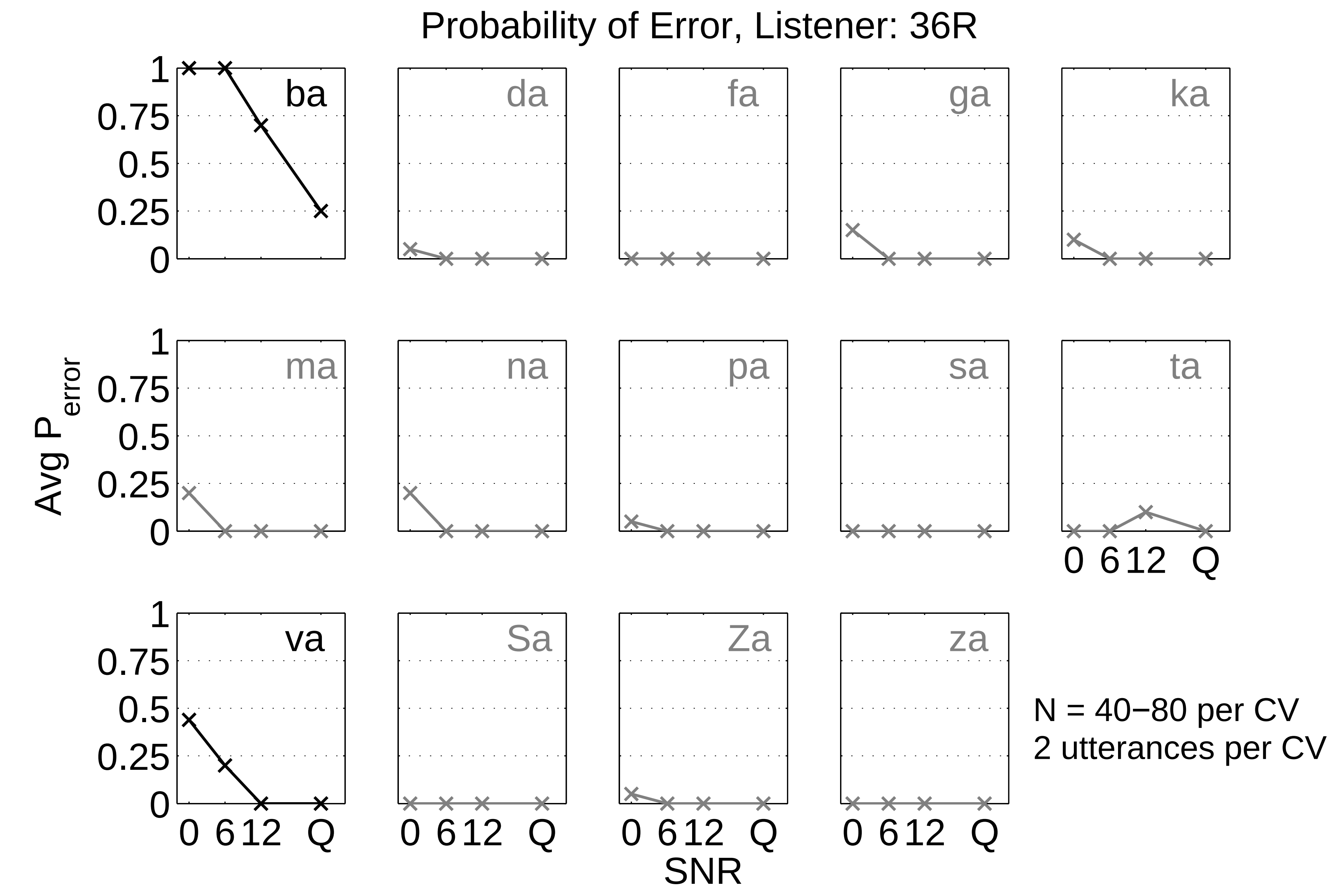
- Why? . . . Too much averaging!

PTA & Average CV error: 36R

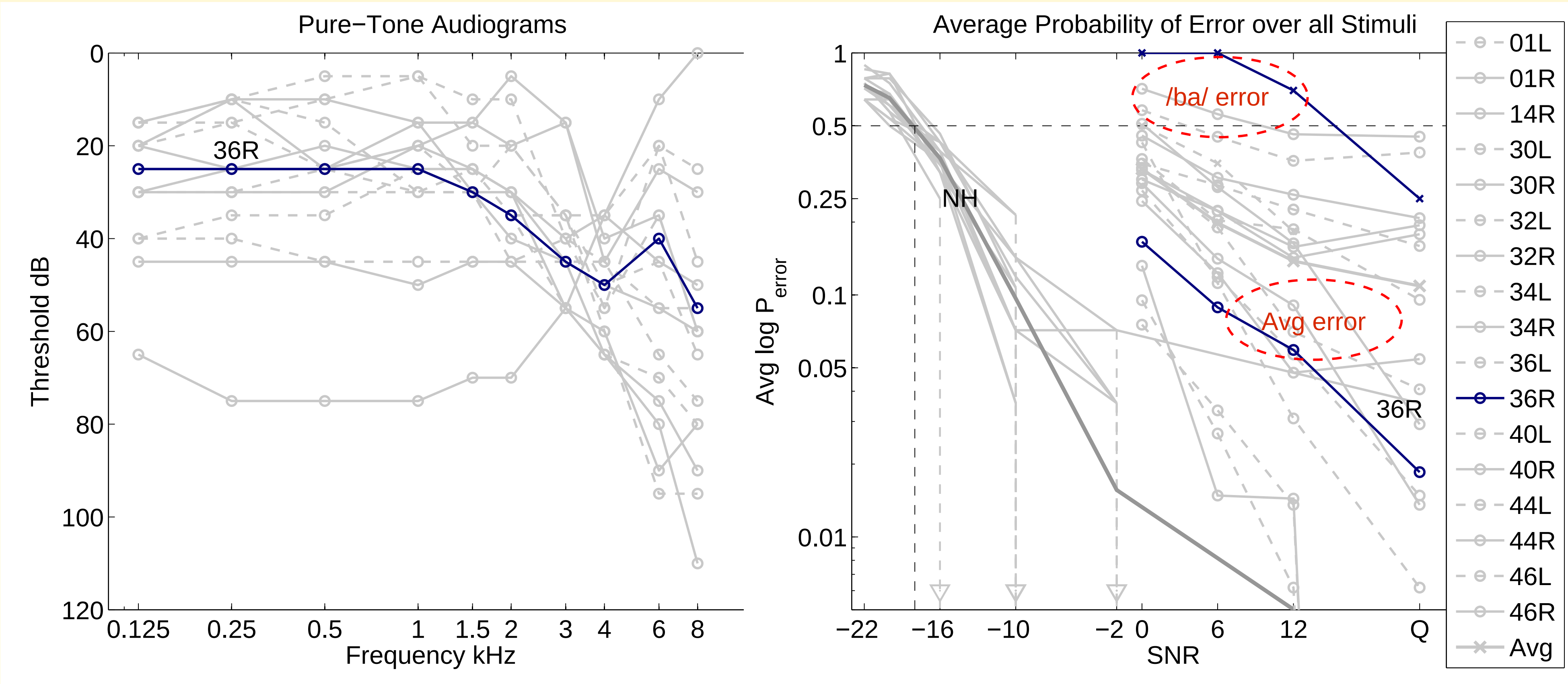
- Case study for 36-R



For Subj 36R: only /ba, va/ have errors!

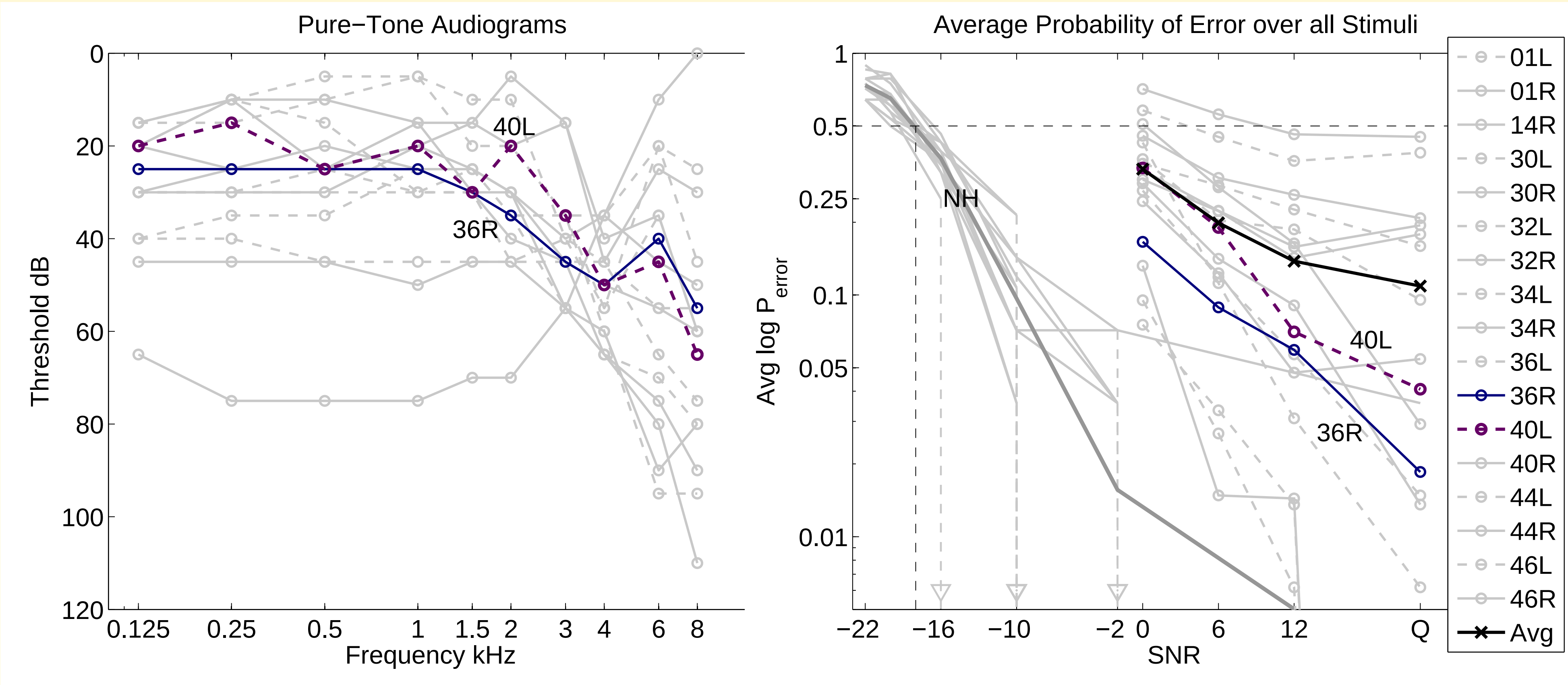


- Case study for 36R



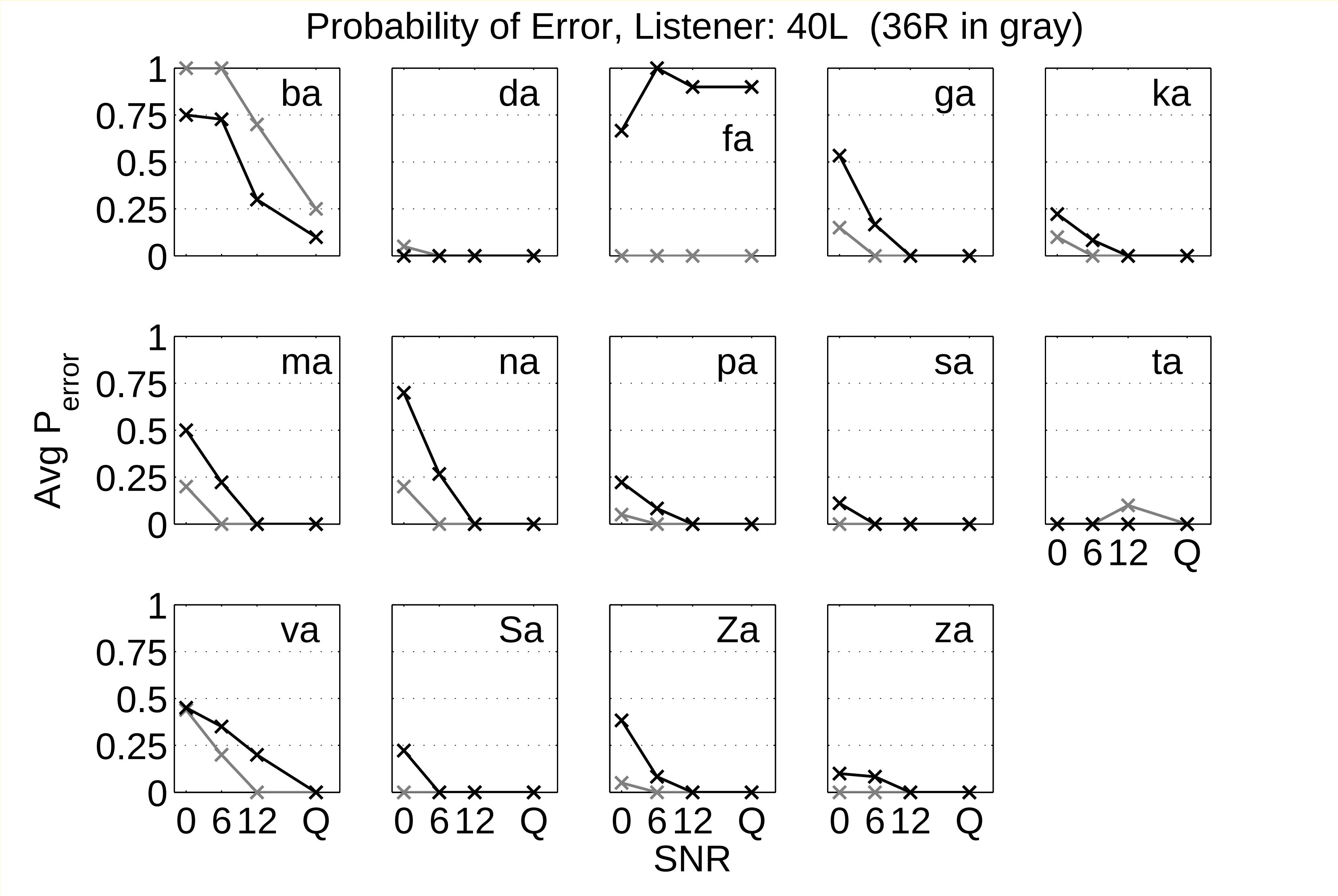
PTA & Average CV error: 40L vs 36R

- Case study for 40L

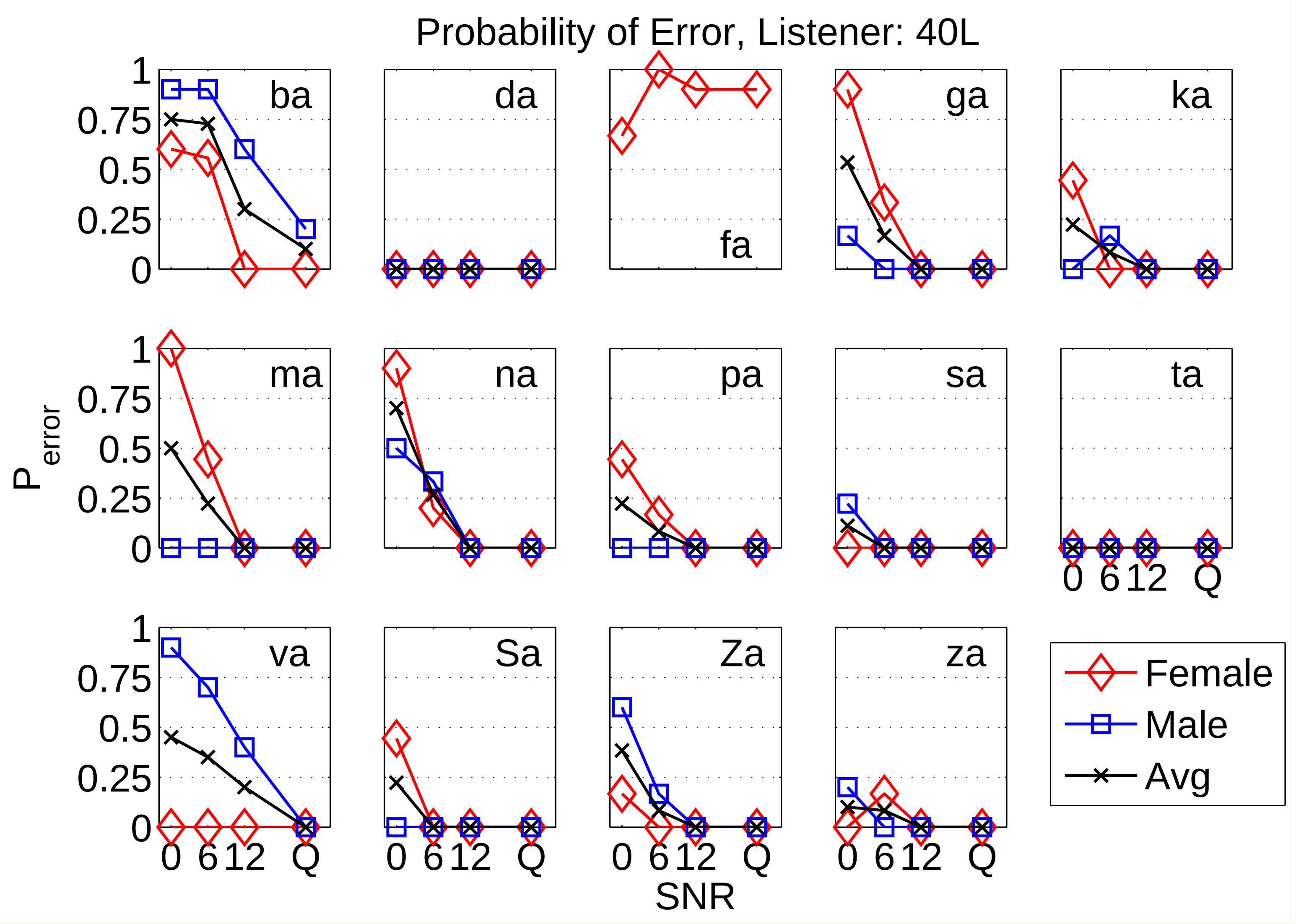


Subject 40L error > 36R

- Errors for 40L re 36R: /ba, fa, ga, ka, ma, na, pa, va, ʒa/

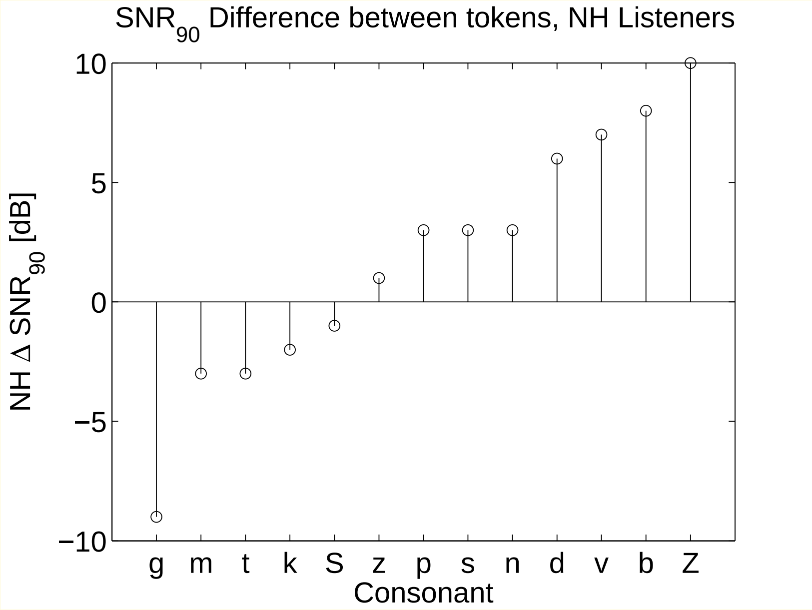


- The token (i.e., talker) errors across CVs are asymmetric



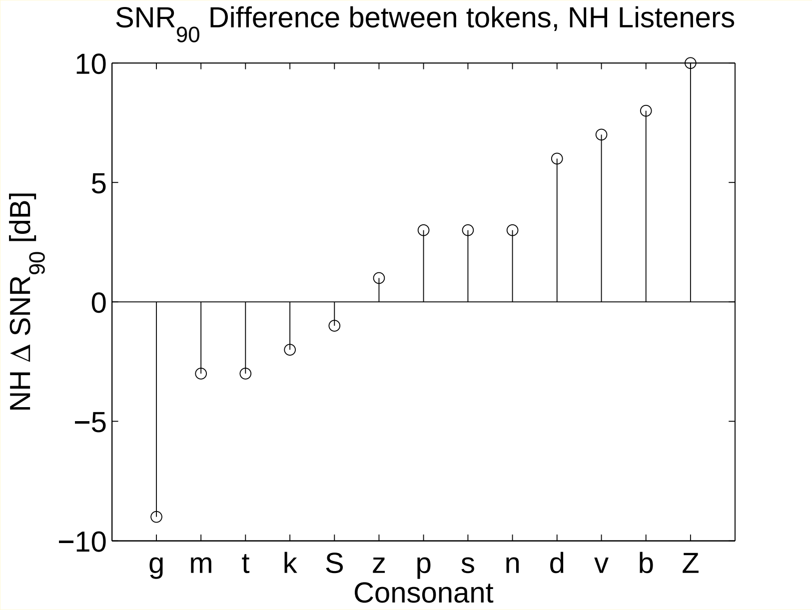
NH vs. HI Utterance differences

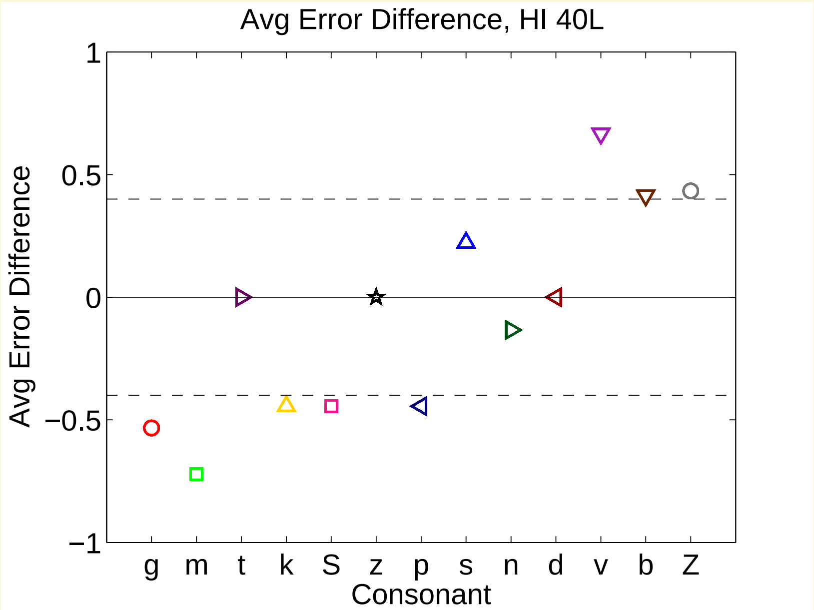
- Rank order of consonants by SNR_{90}



NH vs. HI Utterance differences

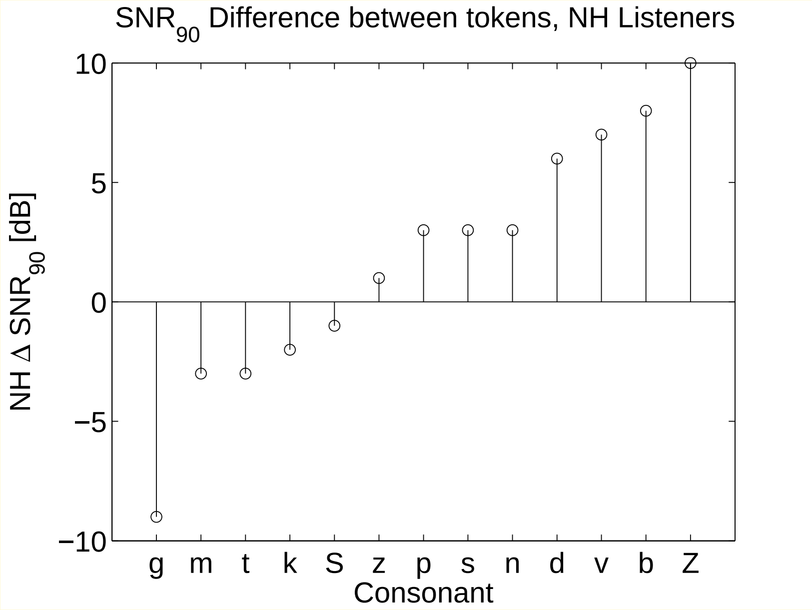
- Rank order of consonants by SNR_{90}

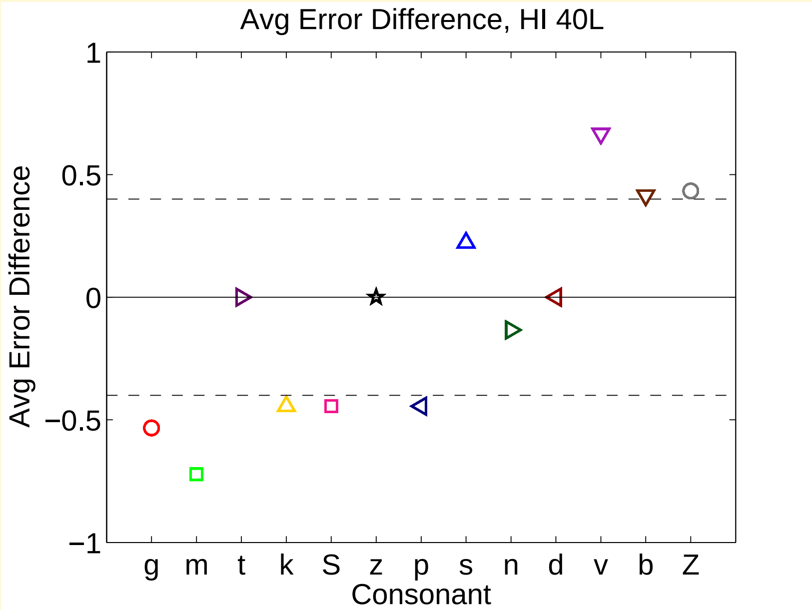




NH vs. HI Utterance differences

- Rank order of consonants by SNR_{90}

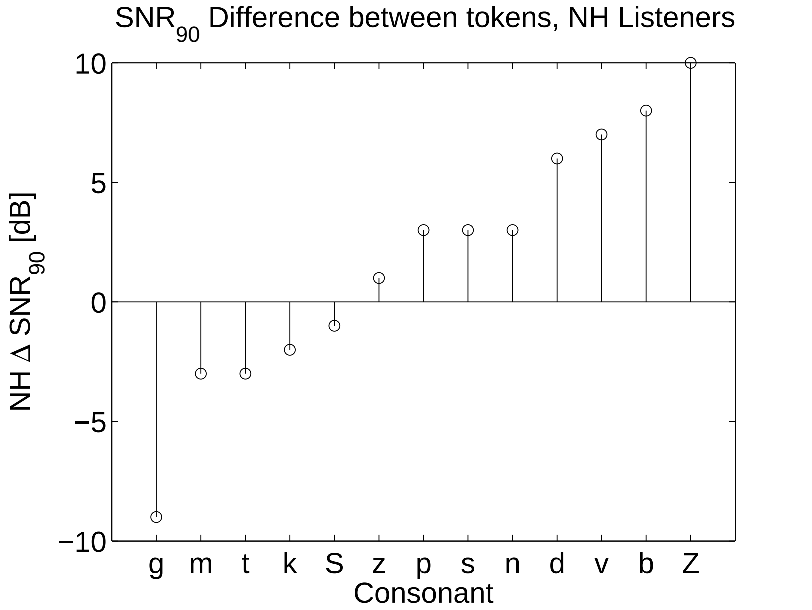


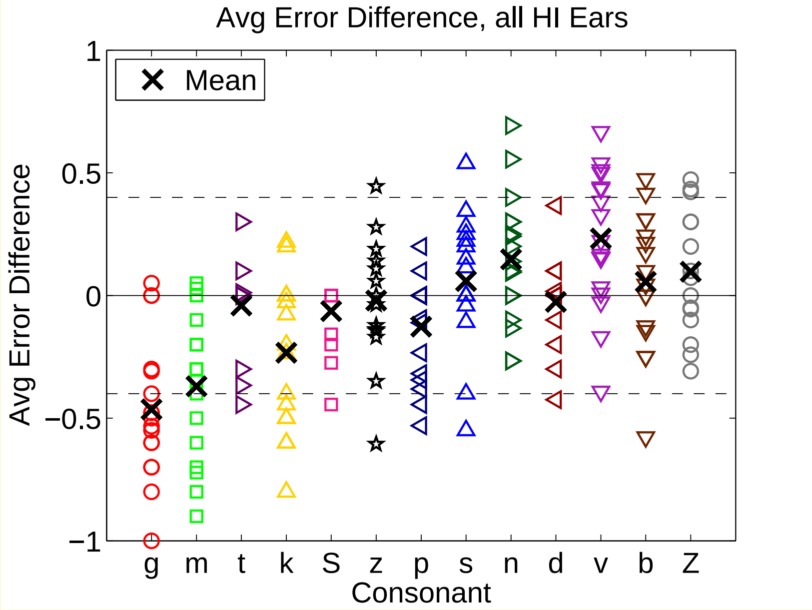


- Note tight relationship in the error difference for NH ears (left) and subject 40L (right)

NH vs. HI Utterance differences

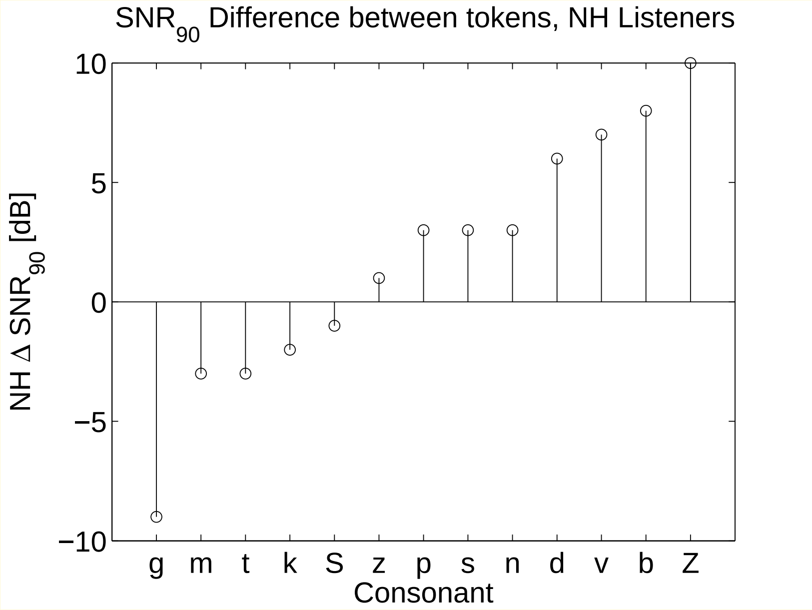
- Rank order of consonants by SNR_{90}

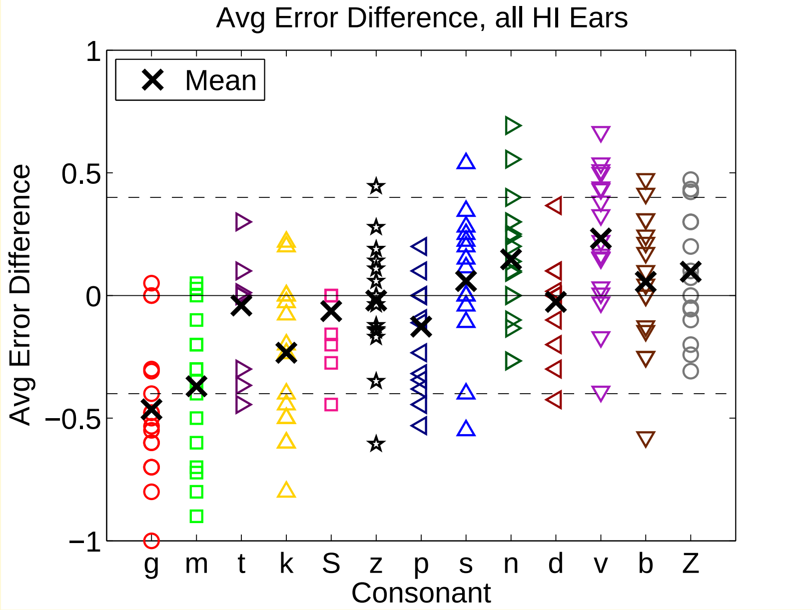




NH vs. HI Utterance differences

- Rank order of consonants by SNR_{90}





- Note the order correlation between NH and HI listeners ($\rho = 0.81$, p< 0.001)

Yogi Berra Quote:

- We made too many wrong mistakes.

Yogi Berra Quote:

- We made too many wrong mistakes.

- The Audiogram and Average Consonant errors have low correlation

Conclusion on HI errors

- The Audiogram and Average Consonant errors have low correlation

Conclusion on HI errors

- The Audiogram and Average Consonant errors have low correlation
- Tokens of a given consonant can have different
  - Errors
  - Confusions

Conclusion on HI errors

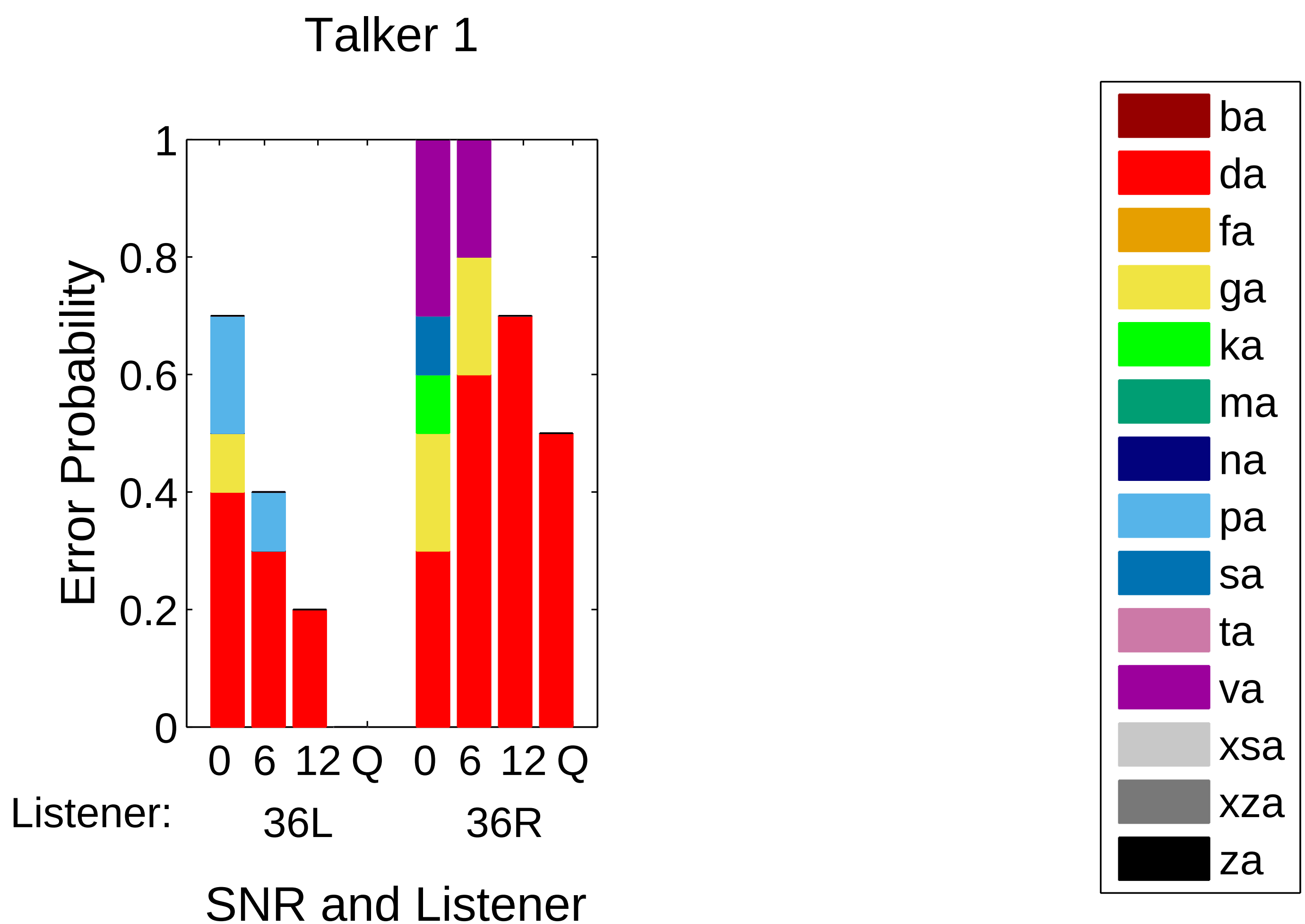
- The Audiogram and Average Consonant errors have low correlation
- Tokens of a given consonant can have different
  - Errors
  - Confusions
- Averaging destroys important HI subject information
  - Making the HI subject *appear* as inconsistent

Conclusion on HI errors

- The Audiogram and Average Consonant errors have low correlation
- Tokens of a given consonant can have different
  - Errors
  - Confusions
- Averaging destroys important HI subject information
  - Making the HI subject *appear* as inconsistent
- The 3 deadly SINs of averaging:
  1. Across *HI subjects:* SIN_s
  2. Across *consonants:* SIN_c
  3. Across *tokens:* SIN_t

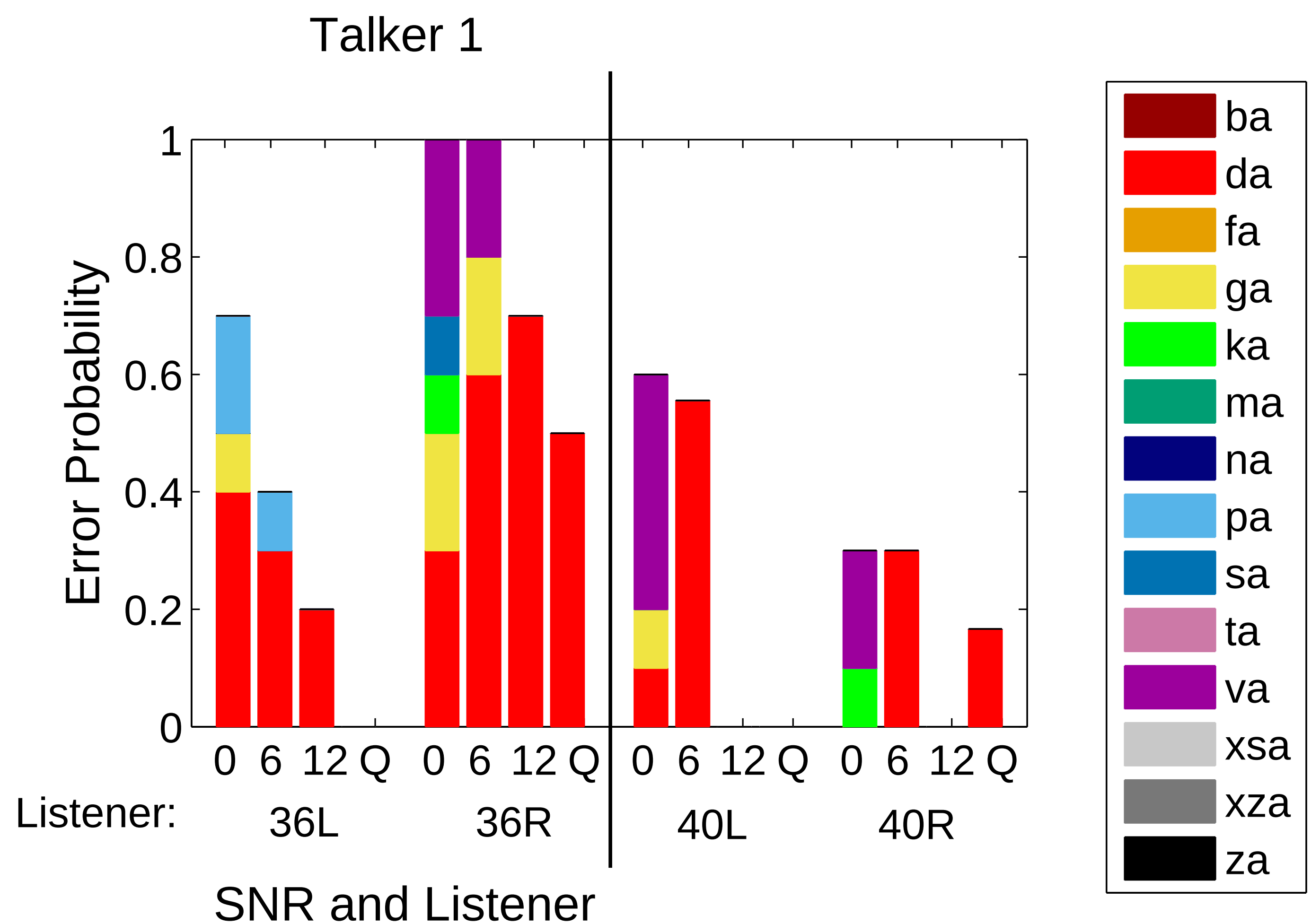
Confusions for Talker 1 of /ba/

- Colors label the confusions



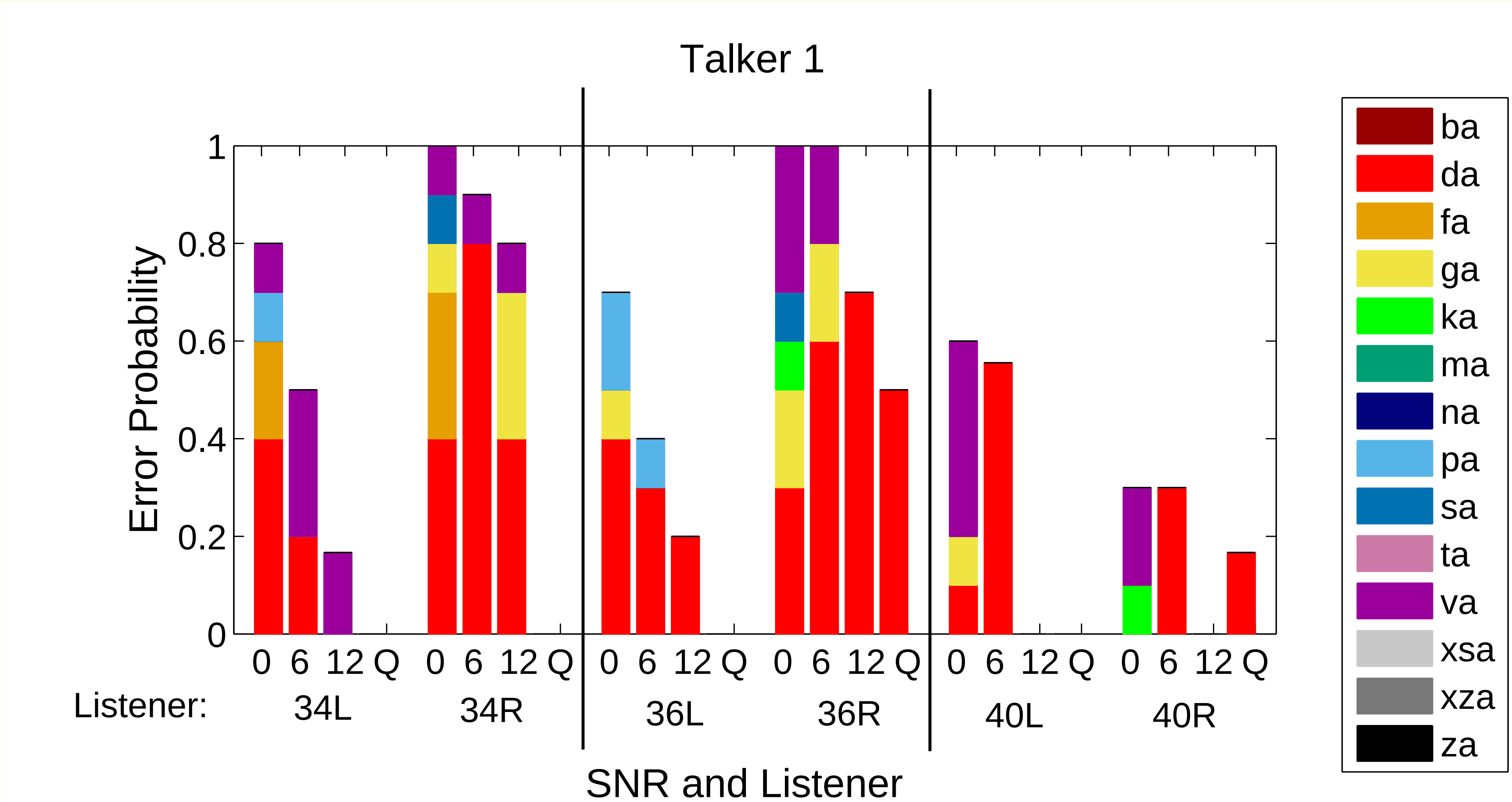
Confusions for Talker 1 of /ba/

- Colors label the confusions



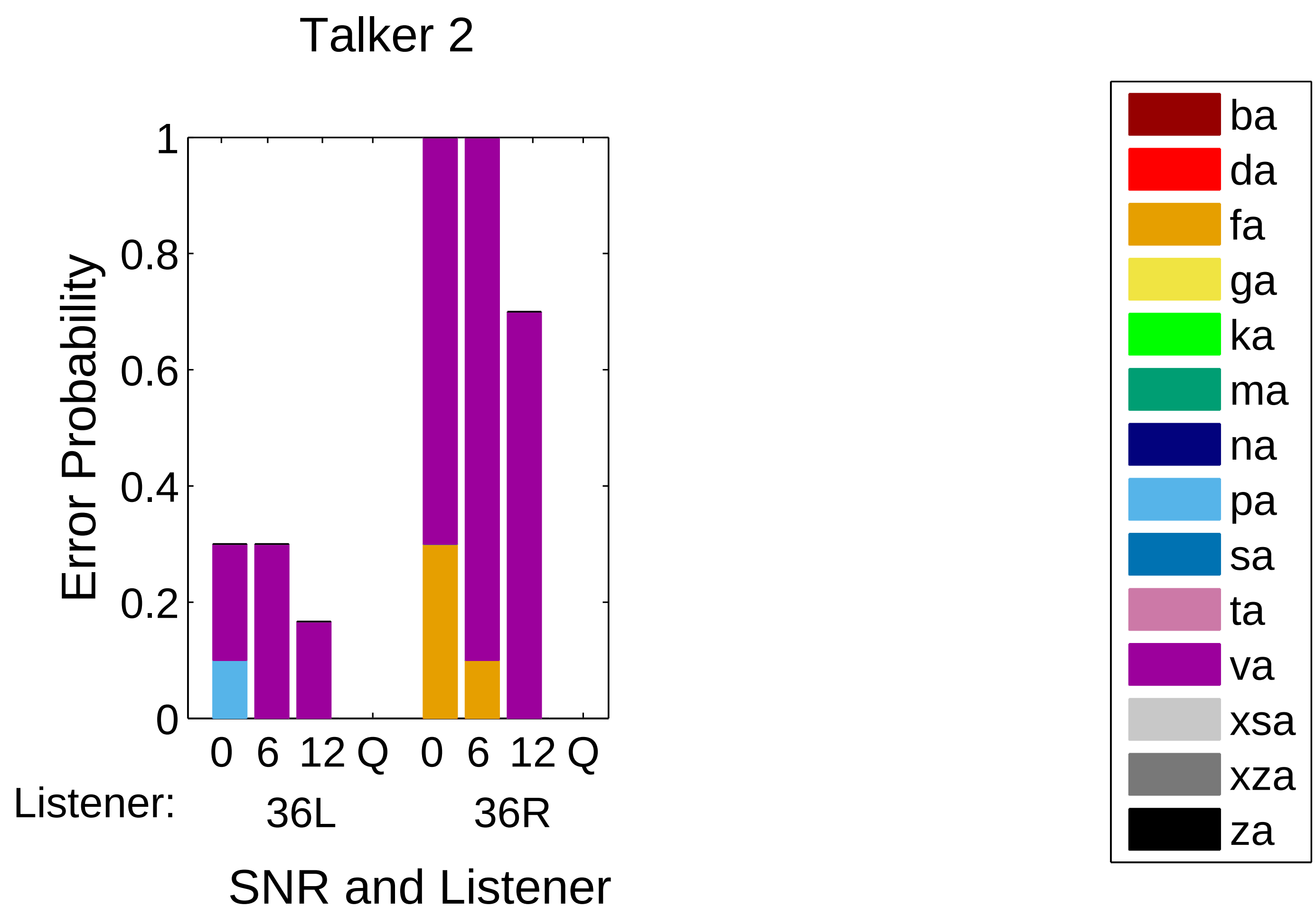
Confusions for Talker 1 of /ba/

- Colors label the confusions



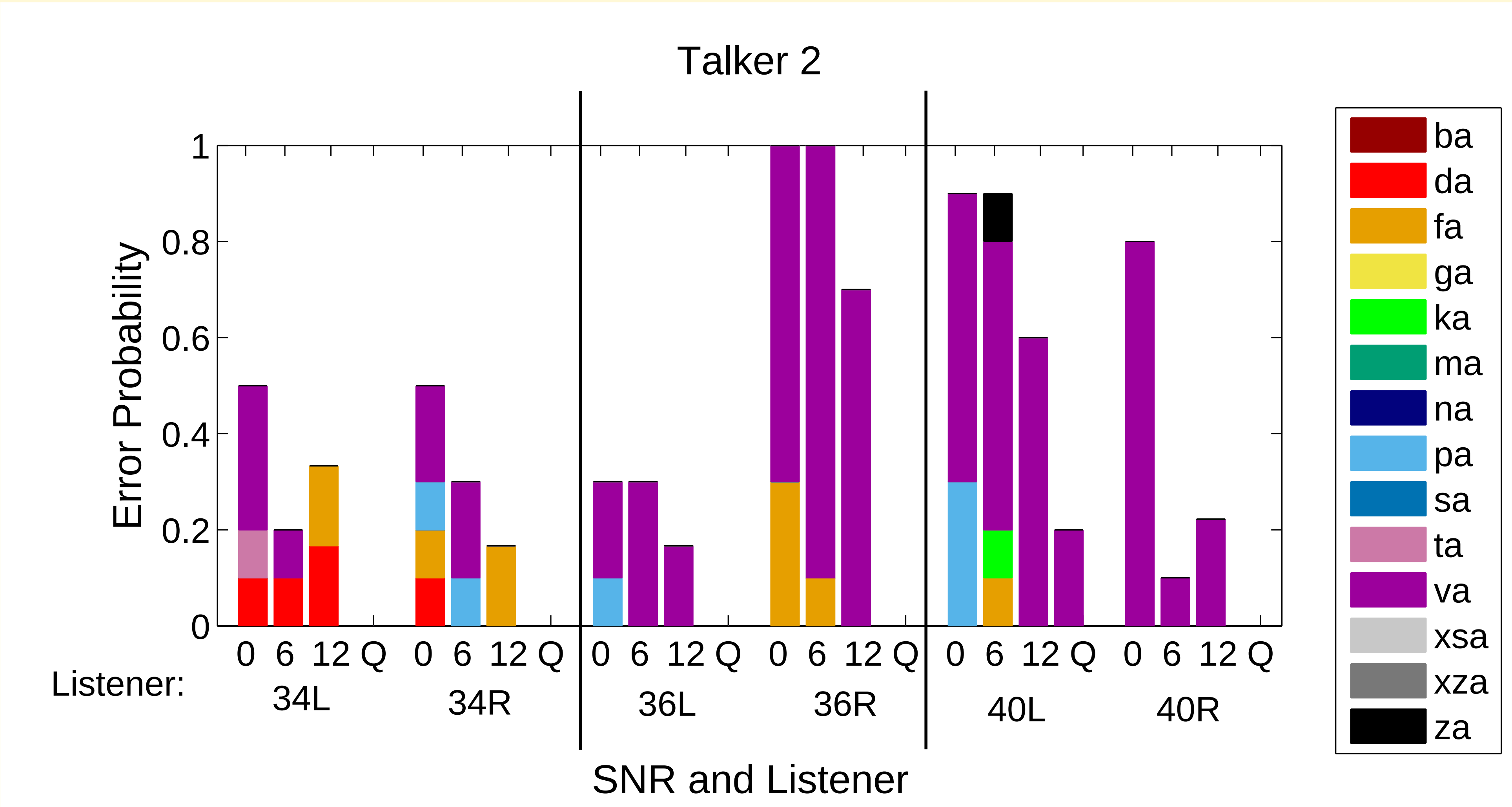
Confusions for Talker 2 of /ba/

- Colors label the confusions



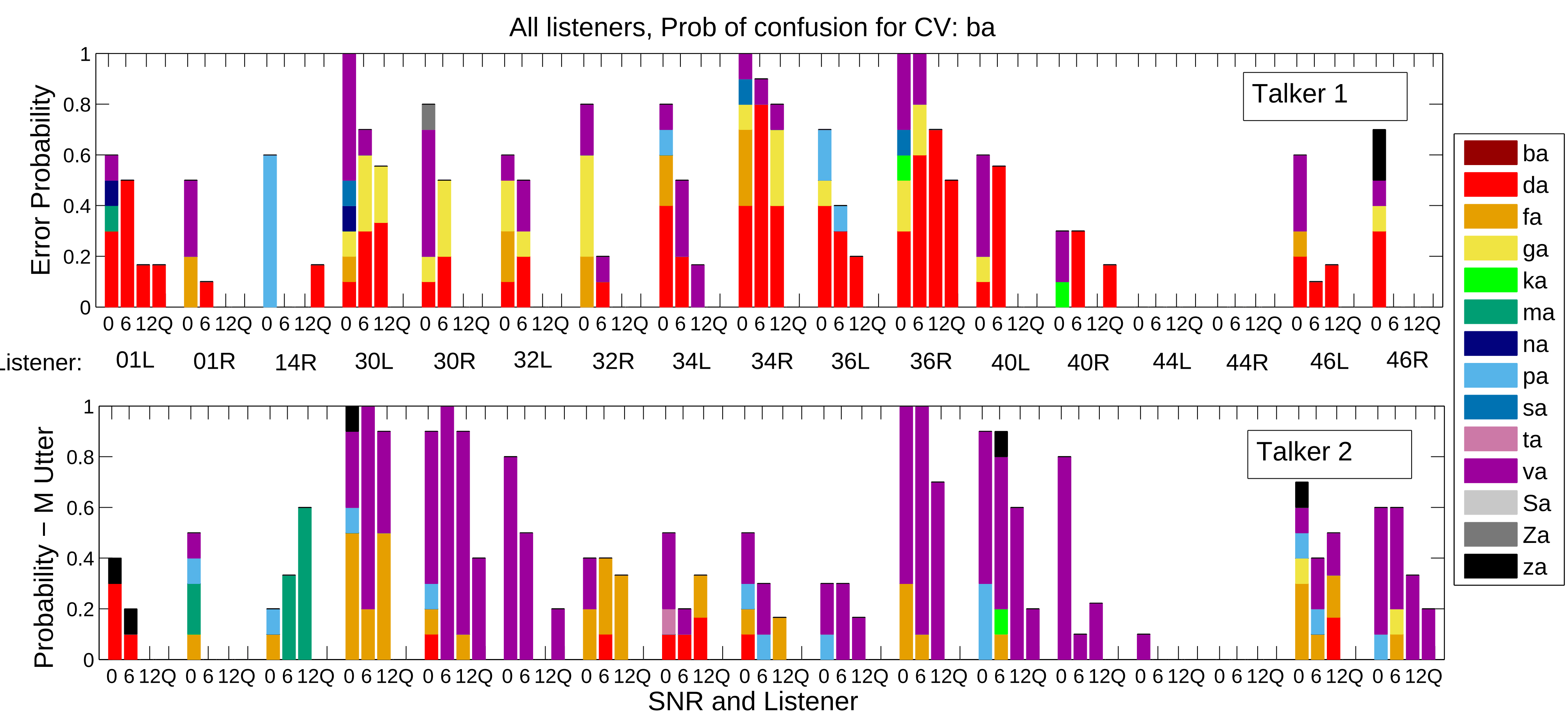
Confusions for Talker 2 of /ba/

- Colors label the confusions



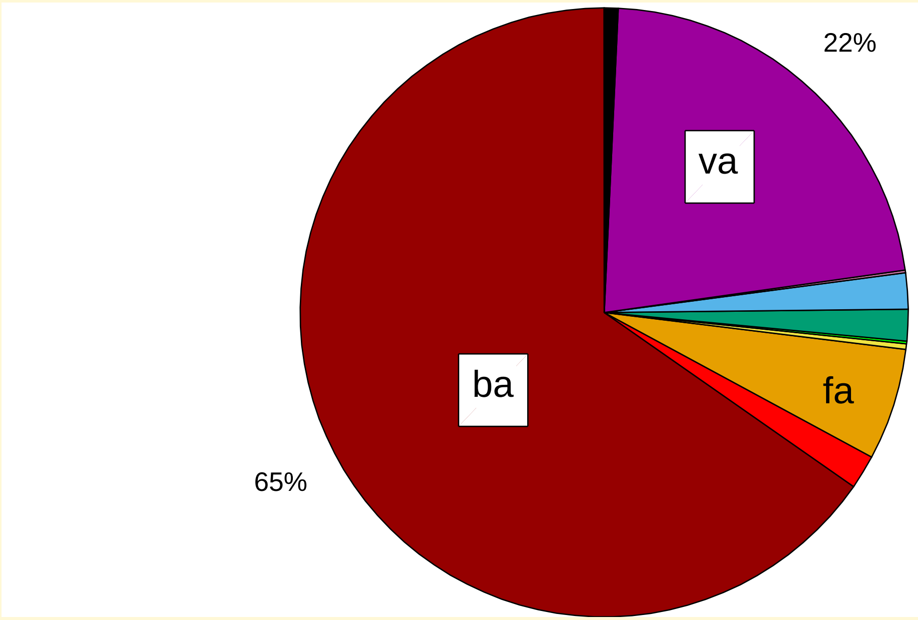
Confusions for two tokens of /ba/

- Colors label the confusions
  - Talker 1 /ba/ → /da/
  - Talker 2 /ba/ → /va/

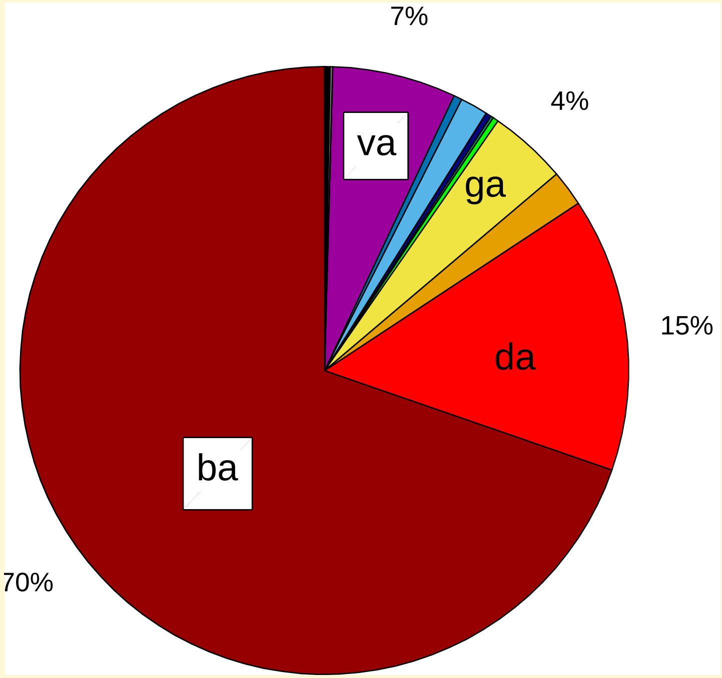


Confusions for two tokens of /ba/

- Average confusions for /ba/
  - Do NOT average across tokens: SIN$_t$ #3 !



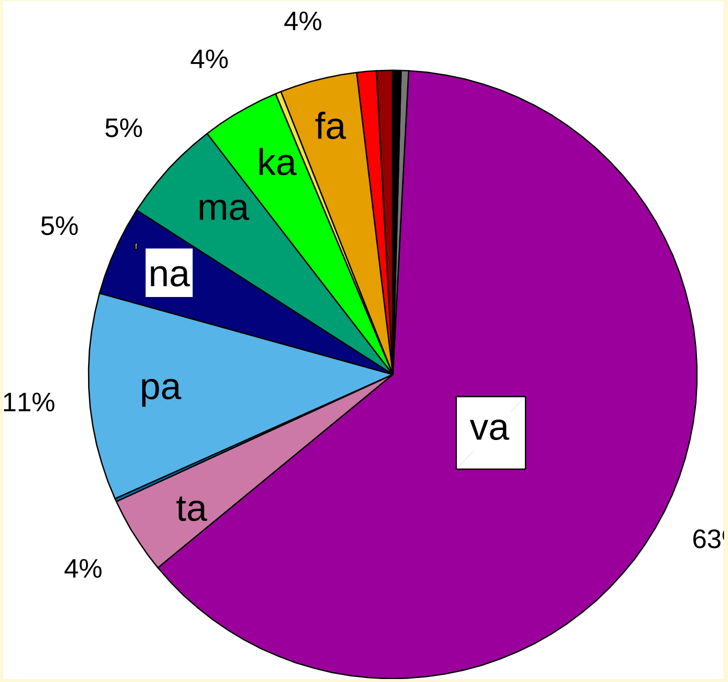
(a) Talker 1

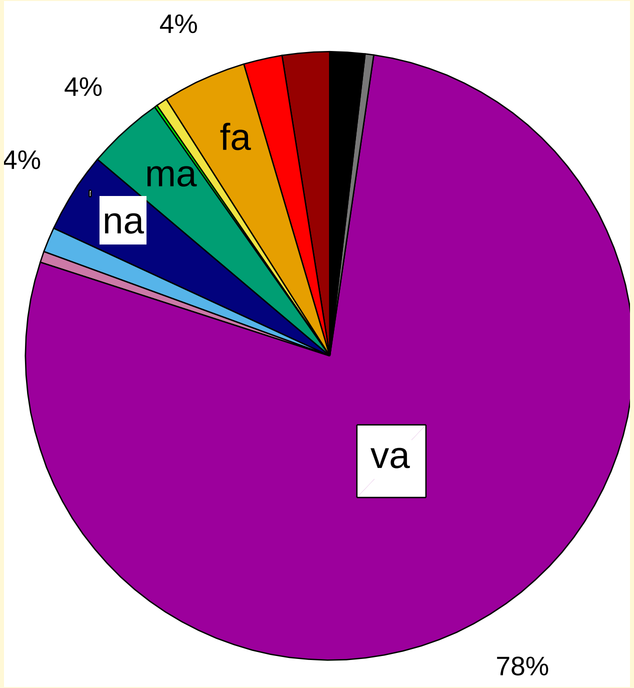


(b) Talker 2

Confusions for two tokens of /va/

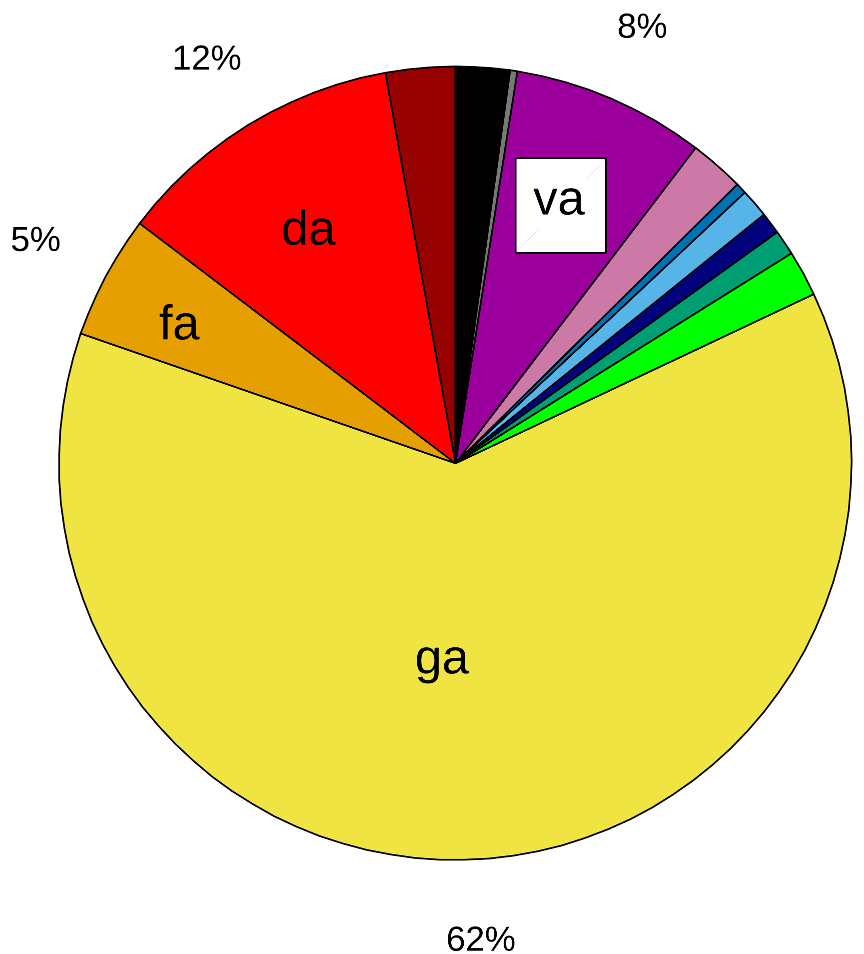
- Average confusions for /va/
  - Do NOT average across tokens: SIN_t #3 !

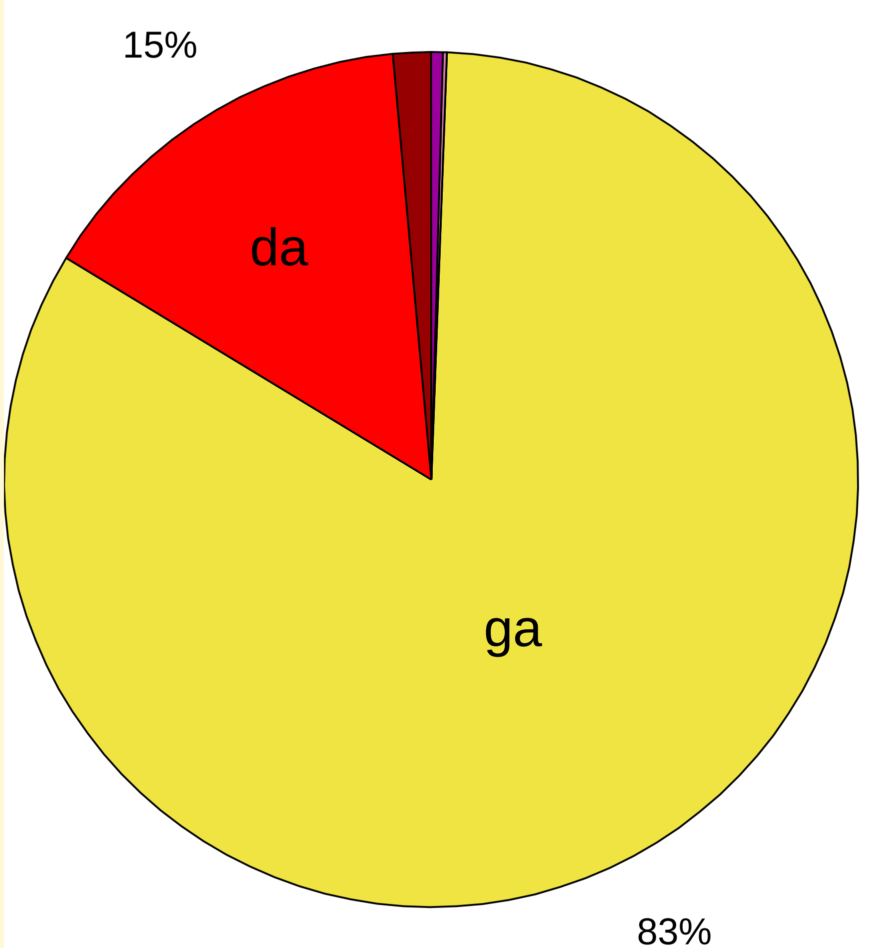




Confusions for two tokens of /ga/

- Average confusions for /ga/
  - Do NOT average across tokens: SIN$_t$ #3 !





Outline

1. Intro + Objectives 2 mins
   - Research objectives
2. Historical overview 4 mins $\Sigma 6$
   - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
   - Speech-feature studies ($>$1950)
3. Methods 8 mins $\Sigma 14$
   - Channel capacity and the Articulation Index
   - Psychophysics of speech/AIgram/3DDS
4. Results with NH ears 10 mins $\Sigma 24$
   - Binary features; Confusions; Primes and Morphs;
5. Results with HI ears 20 mins $\Sigma 44$
   - Individual differences of consonant confusions
6. Summary + Conclusions 6 mins $\Sigma 50$

Conclusion I

- New insights:
  1. AI-gram to visualize speech features

- New insights:
  1. AI-gram to visualize speech features
  2. 3DDS (truncate: time, freq, intensity) to isolated cues: Plosives /p, t, k/, /b, d, g/ + Fricatives /θ, ʃ, ʧ, s, h, f/, /z, ʒ, v, ð/) + vowels /o, e, ɪ/

- New insights:
  1. AI-gram to visualize speech features
  2. 3DDS (truncate: time, freq, intensity) to isolated cues: Plosives /p, t, k/, /b, d, g/ + Fricatives /θ, ʃ, ʧ, s, h, f/, /z, ʒ, v, ð/) + vowels /o, e, ɪ/
  3. Data on discriminating consonants in noise, NH listeners use
     - Plosives: *Burst + timing to Voicing*
     - Fricatives: *Low-frequency edge + duration +* F_0 *modulation*

- We have demonstrated:

1. Speech cue detection is binary (6 dB SNR range)

- We have demonstrated:

1. Speech cue detection is binary (6 dB SNR range)
2. How the AI works:
   - Burst, frequency-edge, timing & SNR_{50} distributions
   - $P_e(\text{SNR}) = e_{\min}^{\textit{SNR}}$ due to SNR^*_{50} distribution

- Findings re HI ears:

1. HI ears have huge individual differences

Conclusion III

- Findings re HI ears:

1. HI ears have huge individual differences
    - Individual differences dominate HI results

Conclusion III

- Findings re HI ears:

1. HI ears have huge individual differences
    - Individual differences dominate HI results
    - No two ears are the same

Conclusion III

- Findings re HI ears:

1. HI ears have huge individual differences
    - Individual differences dominate HI results
    - No two ears are the same
    - Low correlations between HL(f) and $P_e(\mathit{SNR})$

Conclusion III

- Findings re HI ears:

1. HI ears have huge individual differences
   - Individual differences dominate HI results
   - No two ears are the same
   - Low correlations between HL(f) and $P_e(SNR)$
2. Each ear has a different consonant recognition strategy

Conclusion III

- Findings re HI ears:

1. HI ears have huge individual differences
   - Individual differences dominate HI results
   - No two ears are the same
   - Low correlations between HL(f) and $P_e(\mathsf{SNR})$
2. Each ear has a different consonant recognition strategy
3. A better understanding of HI acoustic cue detection will lead to:
   - Improved understanding of HSR for NH & HI ears
   - Better signal processing methods
   - Speech-aware hearing aids in 5 years $>$c2016
     - Fittings based on confusions

Question your basic assumptions

Thank you for your attention

http://hear.ai.uiuc.edu/

http://hear.ai.uiuc.edu/wiki/Main/Publications