Individualizing Noise Reduction in Hearing Aids

Tjeerd Dijkstra, Adriana Birlutiu,
Perry Groot and Tom Heskes

Radboud University Nijmegen

Rolph Houben

Academic Medical Center Amsterdam

Why optimize noise reduction for an individual?

- Most hearing aids have a form of noise reduction
- One of the most cited reasons for not using a hearing aid is “they do not work well in noise” (Kochkin, 2007)
- In current practice noise reduction is either on or off, i.e. none of the parameters of noise reduction can be controlled by a hearing aid dispenser
- In speech quality judgments, hearing impaired show an individual difference unlike the normal hearing who are similar in their ratings (more next).

Speech quality metrics

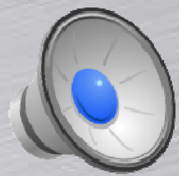
- There is a large body of literature on speech quality metrics, originating from the telephone industry. Overviews are the books by Quackenbush, Barnwell and Clements 1988 and Loizou 2007.
- Examples of metrics are PESQ (Beerends), PEMO-Q (Kollmeier) and Q3 (Kates).
- The performance of a metric is measured by correlating the metric with the **mean** opinion score (MOS) of a group of **trained** listeners.
- How well does a mean-validated speech quality metric predict quality for naive individuals?

Exp 1: speech quality from paired comparisons

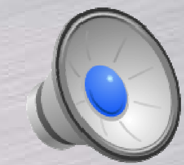
- Data from Arehart, Kates, Anderson & Harvey JASA 2007
- 14 normal hearing and 18 impaired listeners
- 2 HINT sentences as speech material
- 3 distortion conditions: additive noise (from HINT CD), peak clipping and center clipping
- 8 levels for each distortion condition
- Each listener made (3*8)^2 = 576 paired comparisons
- Listeners picked the sound sample that sounded best (least distortion)

Exemplary sound stimuli

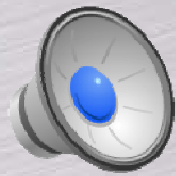
- Additive speech-shaped noise, signal-to-noise ratio of +4 dB



- Peak clipping at 0.1%: highest 99.9% of sound samples are clipped



- Center clipping at 80%: lowest 80% of sound samples are clipped



Pros and cons of Arehart-Kates data set

Cons

- Distortions have no trade-off: the optimal amount of noise is no noise, optimal amount of clipping is no clipping
- No mixtures with some clipping and some noise

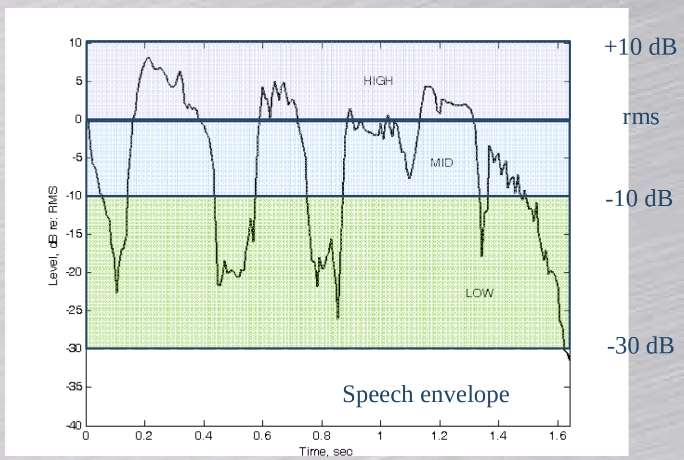
Pros

- Different kinds of distortions compared
- Large data set with 32 listeners and 576 paired comparisons per listener

Kates' speech quality metric

1. Divide clean speech in 3 amplitude regions, low = -30 to -10, mid = -10 to 0, high = 0 to +10 dB re:RMS
2. Calculate coherence for each 20 ms frame and average frames in each amplitude region
3. SpeechPower = Coherence^2 * TotalPower
4. NoisePower = (1 - Coherence^2) * TotalPower
5. Use the standard SII procedure to calculate three SII's, one for each amplitude region

$$Q_3 = 1.73 * CSII_{high} + 2.16 * CSII_{mid} + 2.41 * CSII_{low}$$

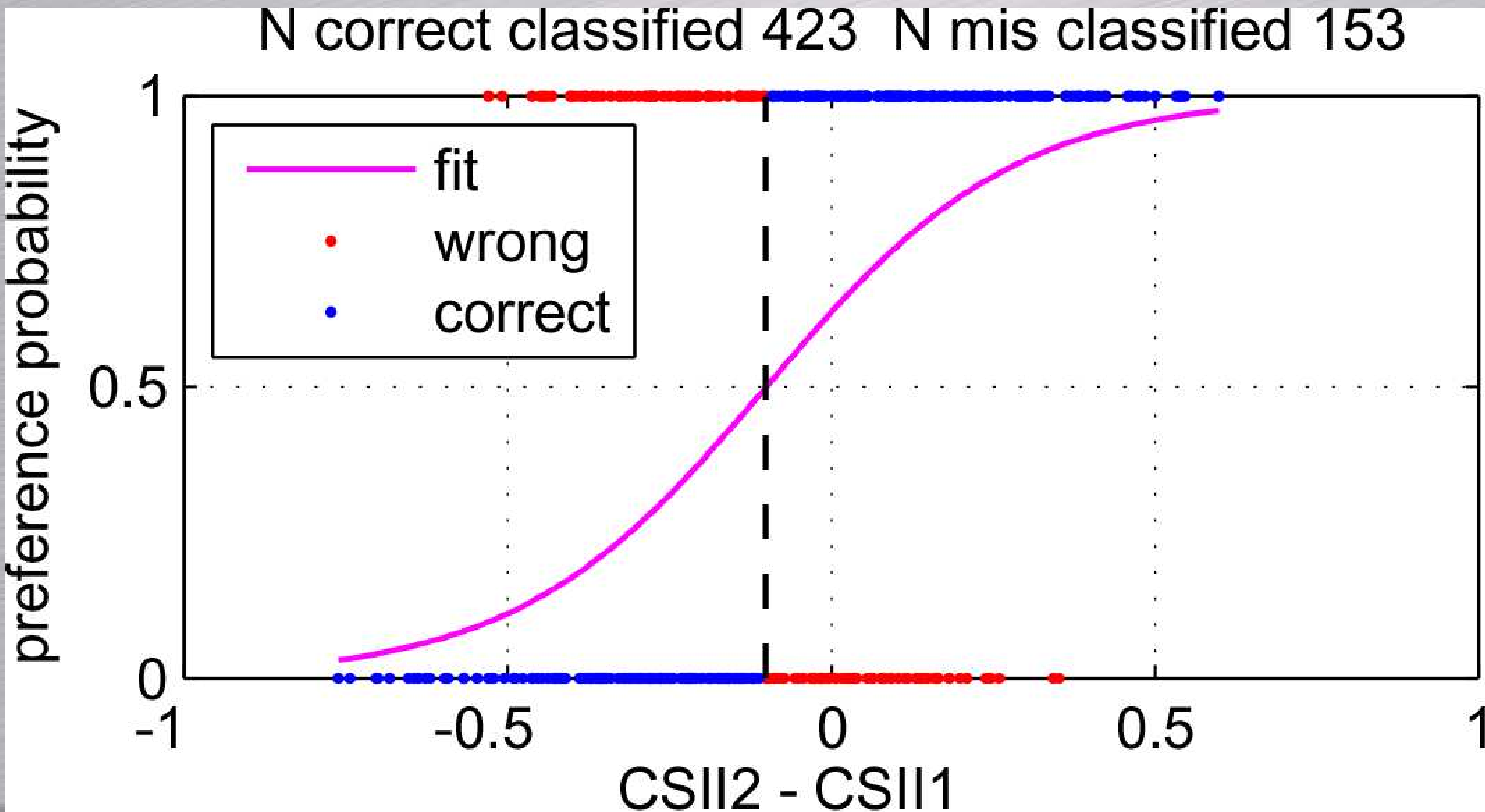


Logistic regression to link preferences to CSII

In words: log odds ~ difference in CSII’s

In equation: log(p/(1-p) = CSII2 – CSII1

Prediction Error = 153/(423+153) = 0.26



Logistic regression results

nh[1-14] normal hearing

hi[1-18] hearing impaired

Q_3 PE = Q_3 prediction error

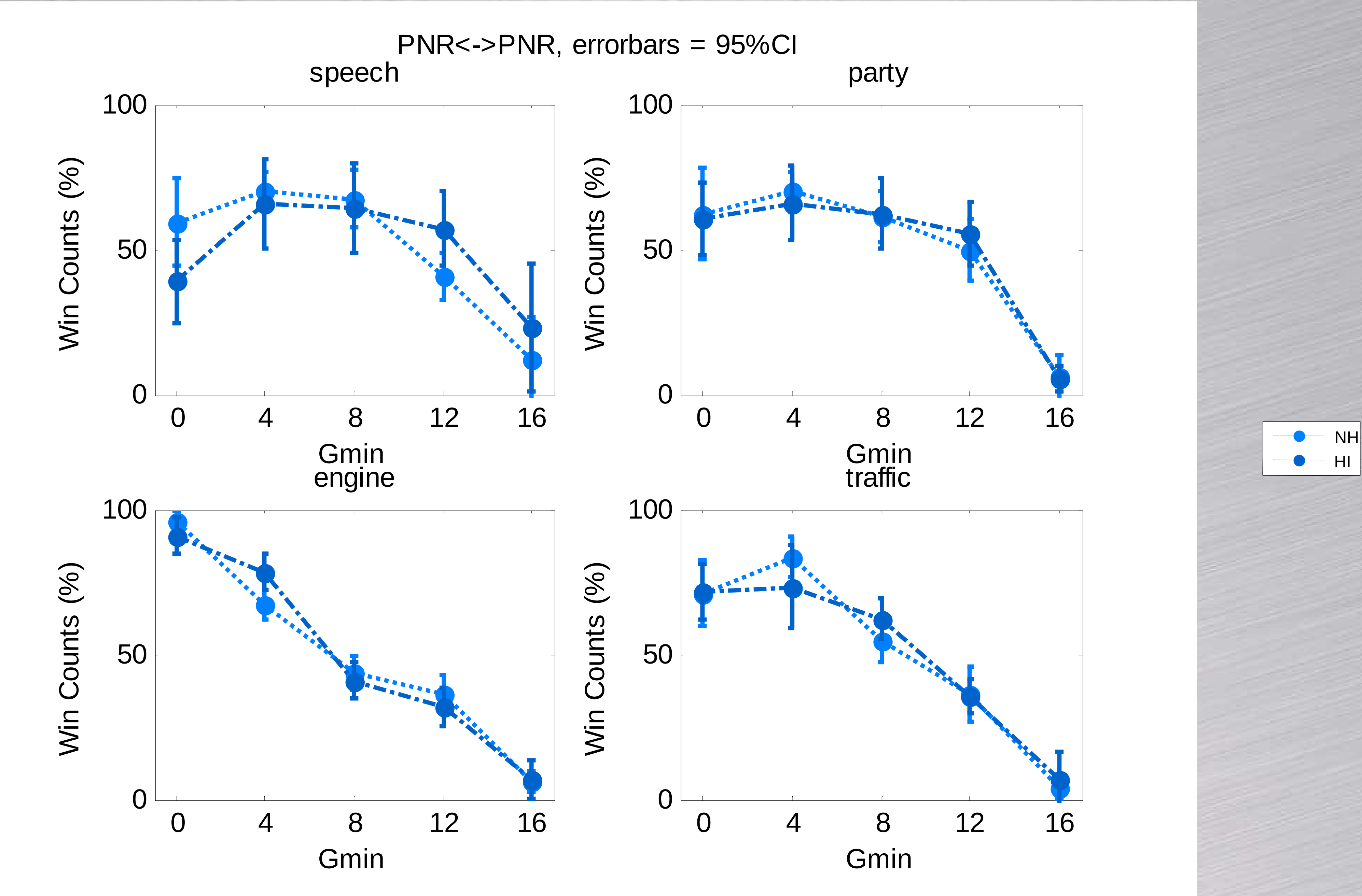
GK PE = individually fitted logistic regression

Subj. name	% bias	% incons.	Q_3 PE	LK PE	GK PE	Q_3 vs LK			Q_3 vs GK			LK vs GK		
						p	s	f	p	s	f	p	s	f
nh1	21.7	0.11	0.15	0.14	0.15	0.81	8	10	0.68	13	10	0.23	8	3
nh2	14.5	0.21	0.13	0.11	0.11	0.14	14	24	0.07	15	28	0.58	5	8
nh3	13.4	0.27	0.12	0.12	0.12	0.86	16	14	0.88	22	20	0.86	16	16
nh4	10.1	0.14	0.12	0.12	0.12	1.00	20	19	0.55	25	20	0.39	8	4
nh5	15.9	0.21	0.14	0.15	0.14	0.15	16	8	0.86	15	15	0.17	9	17
nh6	17.4	0.20	0.14	0.14	0.13	0.86	14	16	0.57	22	27	0.74	16	19
nh7	18.5	0.61	0.18	0.18	0.16	1.00	27	28	0.42	34	42	0.42	24	31
nh8	22.8	1.28	0.21	0.21	0.19	0.88	22	22	0.19	24	35	0.14	17	28
nh9	14.1	0.32	0.16	0.16	0.17	0.87	19	21	1.00	19	18	0.78	26	23
nh10	15.9	0.33	0.14	0.14	0.13	1.00	11	10	0.58	13	17	0.40	9	14
nh11	18.1	0.68	0.16	0.18	0.17	0.05	22	10	0.17	27	17	0.84	11	13
nh12	19.9	0.74	0.15	0.15	0.13	1.00	19	18	0.20	25	36	0.10	17	29
nh13	22.5	0.27	0.15	0.15	0.16	0.87	19	19	1.00	20	19	1.00	5	4
nh14	18.5	0.33	0.14	0.12	0.12	0.16	12	21	0.26	15	23	1.00	6	5
pool	17.4	0.41	0.15	0.15	0.14	1.00	239	240	0.14	289	327	0.07	177	214
hi1	37.7	0.82	0.26	0.19	0.11	0.00	56	94	0.00	29	115	0.00	17	65
hi2	21.7	0.57	0.16	0.11	0.11	0.01	22	46	0.00	20	45	1.00	13	14
hi3	18.8	0.82	0.18	0.16	0.16	0.27	28	38	0.28	37	48	1.00	24	25
hi4	22.8	0.43	0.18	0.15	0.14	0.03	30	51	0.00	24	52	0.32	15	22
hi5	32.2	0.84	0.20	0.14	0.14	0.00	33	64	0.00	33	64	0.80	8	8
hi6	27.9	1.43	0.24	0.23	0.21	0.34	31	40	0.06	37	56	0.22	22	32
hi7	18.1	0.44	0.14	0.11	0.11	0.04	20	36	0.02	17	35	0.79	6	8
hi8	21.7	1.23	0.18	0.17	0.15	0.88	22	24	0.11	22	35	0.14	17	28
hi9	18.1	0.59	0.15	0.12	0.12	0.03	15	31	0.01	15	35	0.62	16	20
hi10	15.6	0.26	0.14	0.12	0.11	0.01	7	21	0.01	12	31	0.42	10	15
hi11	29.3	0.22	0.16	0.12	0.10	0.01	22	44	0.00	18	53	0.03	9	22
hi12	35.5	0.31	0.19	0.14	0.09	0.00	28	56	0.00	18	79	0.00	15	48
hi13	35.9	2.00	0.24	0.22	0.18	0.27	44	56	0.00	45	78	0.00	15	36
hi14	25.0	0.66	0.18	0.13	0.11	0.00	21	52	0.00	16	58	0.03	6	17
hi15	14.1	0.22	0.13	0.14	0.12	0.80	32	29	0.72	33	37	0.19	7	14
hi16	30.1	2.05	0.31	0.23	0.20	0.00	63	109	0.00	45	108	0.06	27	44
hi17	31.9	3.04	0.24	0.22	0.22	0.28	37	48	0.19	36	49	0.75	4	6
hi18	30.4	0.49	0.18	0.15	0.15	0.05	19	34	0.04	24	42	0.69	11	14
pool	25.9	0.91	0.19	0.16	0.14	0.00	530	873	0.00	481	1020	0.00	242	438

Exp 2: speech quality of noise reduction algorithms

- Two algorithms, PNR a standard spectral subtraction one and NSNR, a fancier non-stationary noise reduction one
- 10 normal hearing and 7 impaired listeners
- 224 Dutch sentences as speech material
- 4 noise types, stationary speech shaped, babble, car noise and street noise
- 2 presentations (test and one retest)
- 5 levels of noise reduction expressed as the maximal noise reduction gain Gmin = 0, 4, 8, 12, 16 dB
- Each listener made 224 paired comparisons
- Listeners picked the sound sample that sounded best

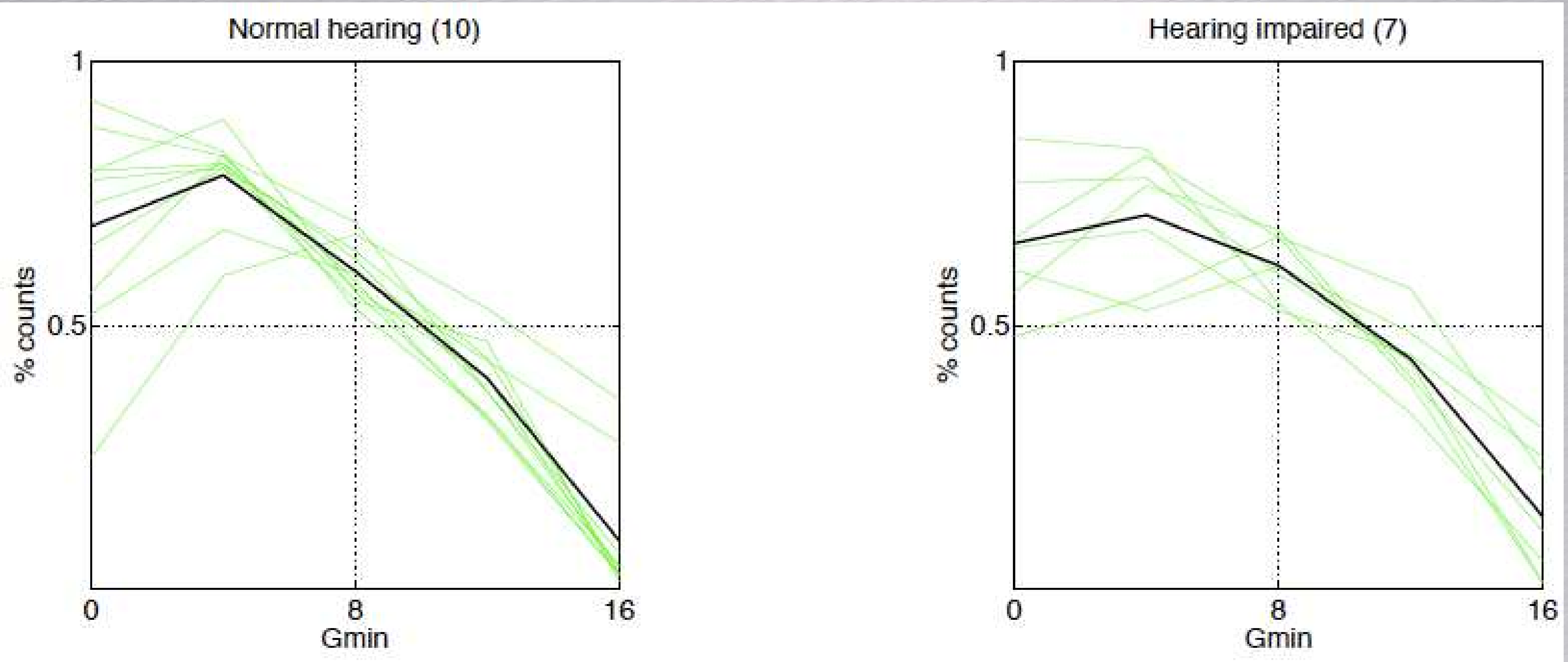
Mean results over listeners



Results of individual listeners

Some listeners prefer Gmin = 0 (noise reduction off), some Gmin = 4 and some Gmin = 8 dB.

-> each listener makes a different trade-off between residual noise and distortion



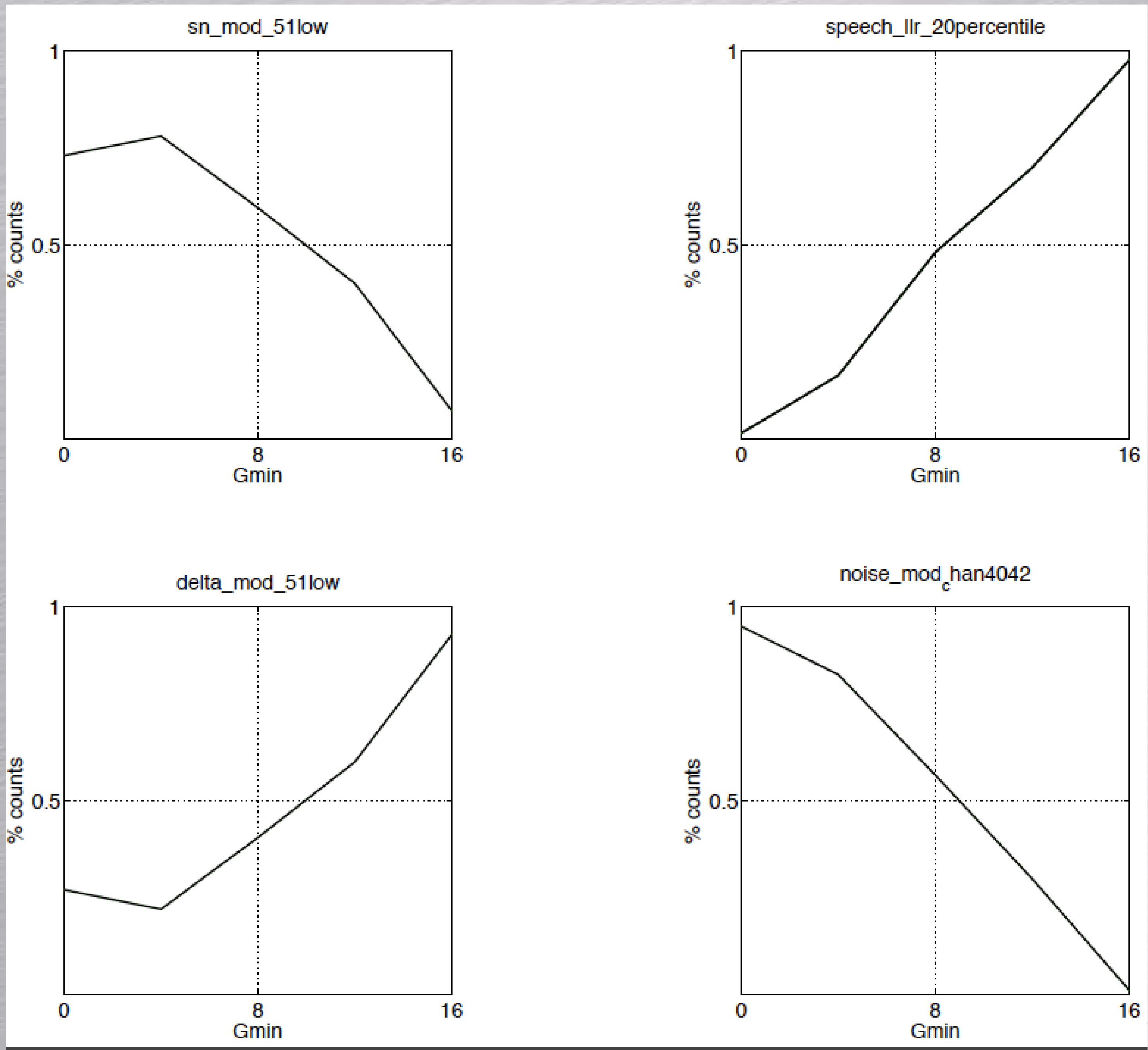
How can we capture noise reduction preferences?

Step 1: we calculated a large set of features, both on the noisy speech, the clean speech and the pure noise, in all about 300 features. Probably everyone's favorite was in our set! Features were calculated from (1) noisy speech (2) speech only (3) noise only (4) difference between noisy and clean speech

Step 2: by pooling all data and with help of a sparse logistic regression we selected the four best features.

Step 3: with individual fits we show that this feature set captures individual behavior.

Best features, L1-regularized logistic regression



prediction accuracy

Set of four features predicts performance about as well as a single individual feature picked from all 300 features, except for listeners with a low test-retest consistency.

Subject	Individual best features	Combined features	Consistency
#1	0.89	0.87	0.79
#2	0.93	0.87	0.61
#3	0.62	0.57	0.59
#4	0.82	0.82	0.71
#5	0.84	0.82	0.84
#6	0.87	0.83	0.76
#7	0.91	0.89	0.84
#8	0.89	0.85	0.87
#9	0.67	0.61	0.56
#10	0.78	0.77	0.82
#11	0.67	0.58	0.52
#12	0.88	0.83	0.82
#13	0.77	0.74	0.74
#14	0.86	0.82	0.78
#15	0.66	0.60	0.64
#16	0.83	0.81	0.79
#17	0.85	0.68	0.58

Conclusions

1. Quality metrics for the hearing impaired need to be personalized, while a single quality metric seems adequate for normal hearing listeners.
2. Listeners show an individual preference for noise reduction parameter values: some are more bothered by the noise (prefer Gmin = 4 or 8 dB) and some more by the speech and noise distortion (prefer Gmin = 0 dB)
3. The long term goal of the HearClip project is to provide algorithms for individual fitting of hearing aids. Sound features that capture individual patterns are an essential component of this approach.

HearClip vision