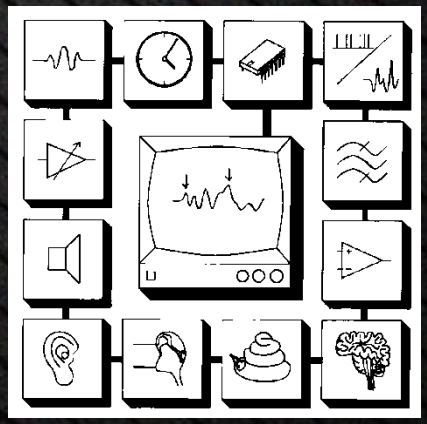
Pre- ANHSC 2015 conference
Audiology workshop

Quality issues in ABR recordings

Guy Lightfoot
ERA Training & Consultancy
Email: admin@eratraining.co.uk
www.eratraining.co.uk www.abrpeerreview.co.uk



Background

- Davis & Bamford report 1997
  - Made the case for universal newborn screening (OAE)
- Preparation of protocols
  - 1999 diagnostic ABR based on click ABR
- National programme roll-out 2001 – 2006
- 3-day ABR training course on ABR for all staff
- Development of suite of protocols (now “guidance”)
  - Early Assessment overview
  - AC & BC frequency specific ABR & ASSR testing
  - ANSD & CM testing
  - Others, inc Tymp & VRA
  - Available at www.abrpeerreview.co.uk/resources.html

Background

- National database (eSP)
- Screen Includes:
  - Bilateral PCHI of moderate or greater degree (>=40dB averaged 0.5 to 4kHz)
  - ANSD in NICU/SCBU babies
- Screen Excludes
  - Unilateral PCHI (but will be detected)
  - Mild bilateral PCHI (some may be detected)
  - ANSD in well babies
- Referred babies should be assessed within 4 weeks
- Assessment to be completed by 8 weeks

Prevalence

PCHI Bilateral – 1.3 per 1000 births

- 1.1 Congenital
- 0.2 Acquired

PCHI Unilateral: - 0.8 per 1000

Progressive – inc CMV

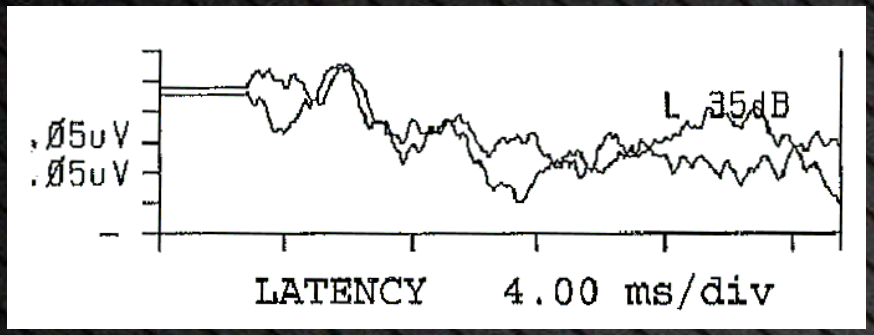
1.65 per 1000 by age 9 yr

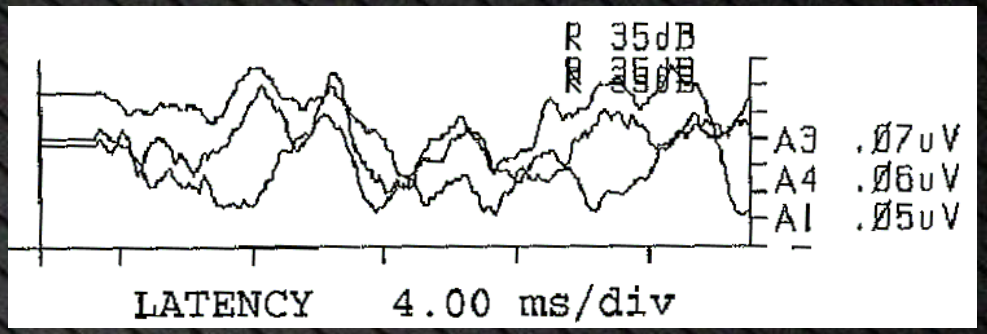
ANSD ~ 0.1 per 1000

ABR Quality issues

- Despite comprehensive prescriptive guidance, several “serious untoward incidents” still occurred
- Out of court settlements typically £1.5M (AU$2.8M)
- Series of QA audits were initiated from 2009
- Audits have revealed:
  - National guidance sometimes ignored or misunderstood
  - Errors of test parameters
  - Errors of waveform interpretation
  - Errors of test strategy
  - Errors of reporting
  - Errors of case management

An example:
Discharged but baby had a profound loss





- Errors:
  - "Auto" display gain (note Rt ear scales)
  - Far too lax an artifact rejection level
  - Tester stopped averaging when they felt a response was probably there (too few sweeps)
  - Very lax interpretation
  - Tester attitude: "I've been doing this for years; I don't need a protocol to tell me what to do"

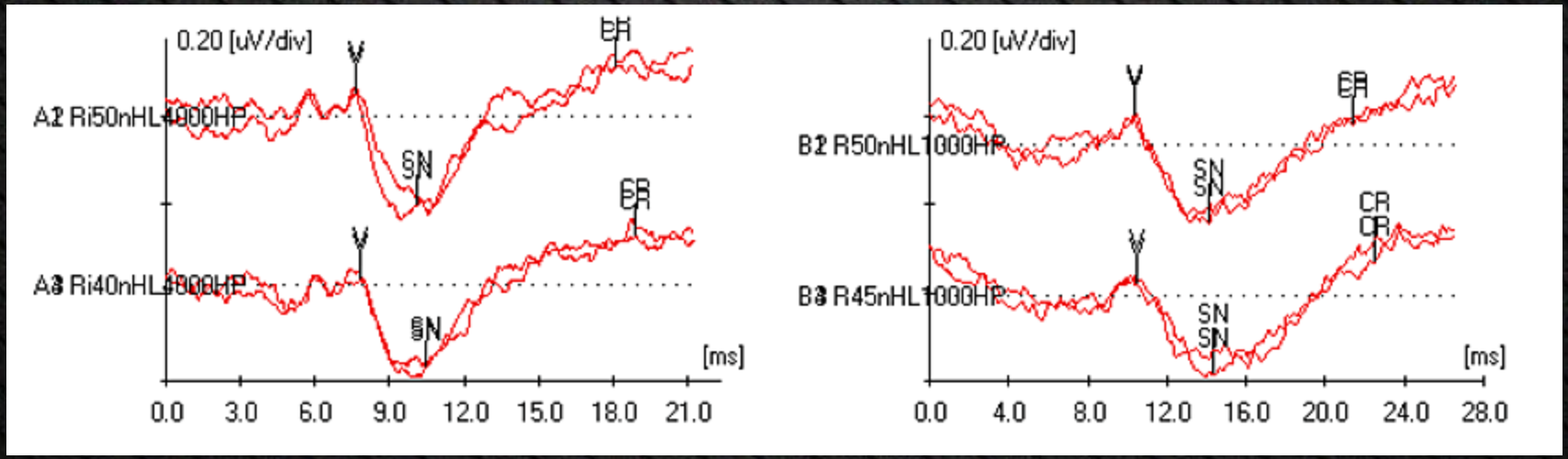
Errors of test parameters

- “Otoneurological” parameters used
  - e.g. 100Hz HPF: attenuates both noise & response, so responses close to threshold may not be recognised
  - e.g. epoch too short so responses close to threshold not recognised (esp low frequency)
- Hazardously high stimulus level with inserts in babies
  - 100dBnHL (clicks) = 120dBnHL in canal = >145dBpeakSPL
- Too few sweeps
  - e.g. <2000: excess noise compromises interpretation
  - now moving toward objective measurements to guide us
- Too lax an artifact rejection level
  - e.g. 15-20µV: excess noise allowed to contaminate the ABR

NHSP ABR Guidance for AR level

- 1999: (clicks) ± 10 - 25µV
- 2001: (tone pips) ± 10 - 15µV
- 2008: ± 5 - 10µV
- 2010: (current) ± 3 - 10µV (default 5µV)

- Example: 4k & 1k; 2000 sweeps, ± 5µV, sleeping baby



But test conditions are not always ideal

- If 5µV creates excessive rejects, what do we do?
  - Wait! Most babies will settle
- What if they don't?

- Options include:
  - Stick with 5µV and suffer +++ rejects (takes longer)
  - Increase to 7 or 10µV and accept noise into the average (but doubling noise requires 4 times as many sweeps – also takes longer)
- Which approach is the most time-efficient?

Study Design

- 26 typical babies referred from newborn screen
- Tested with NHSP recommended parameters
  - 4kHz 5-cycle tone pips at 49.1/s,
    30 / 40dBeHL or at threshold / threshold +10dB
- “EEG” with ±40µV rejection and trigger pulses recorded onto data logger for off-line re-averaging
- 100, 3000-sweep (61.1s) epochs re-averaged using:

  ± 5µV

  ± 6.5µV

  ± 8µV

  ± 10µV (conventional + Bayesian averaging)

  ± 20µV (conventional + Bayesian averaging)

Bayesian Averaging?

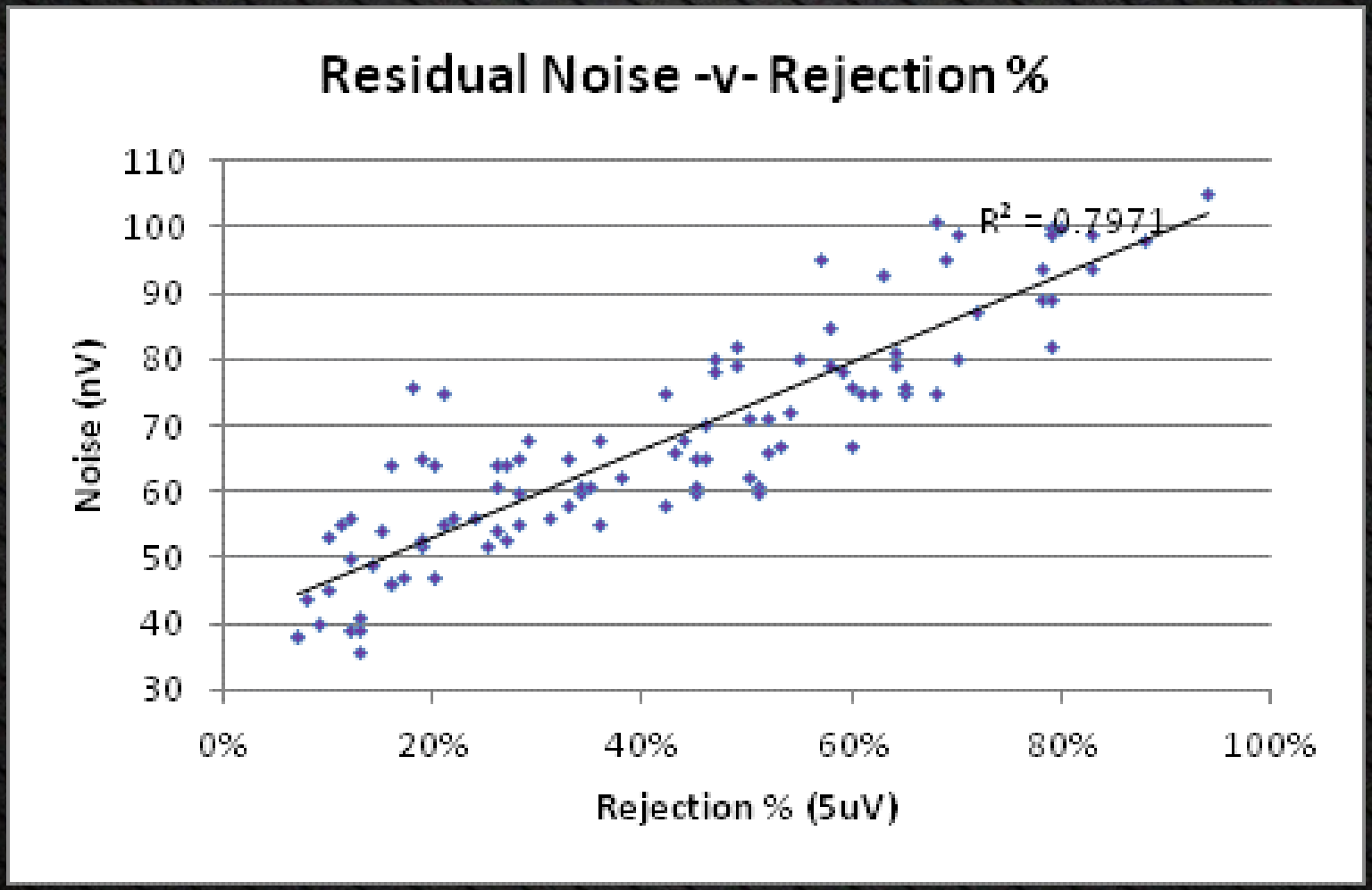
- Adopt a more lax AR level
- Residual noise measured in each 100-sweep block
- Each block is weighted: 1 / residual noise
- Final average computed from weighted blocks
- Advantages:
  - Noisy periods have less destructive effect
  - Average is dominated by periods of lower noise
- Disadvantages:
  - No benefit if noise in each block is similar
  - Regular noise (e.g. cardiac activity) is not rejected

How should we measure “efficiency”?

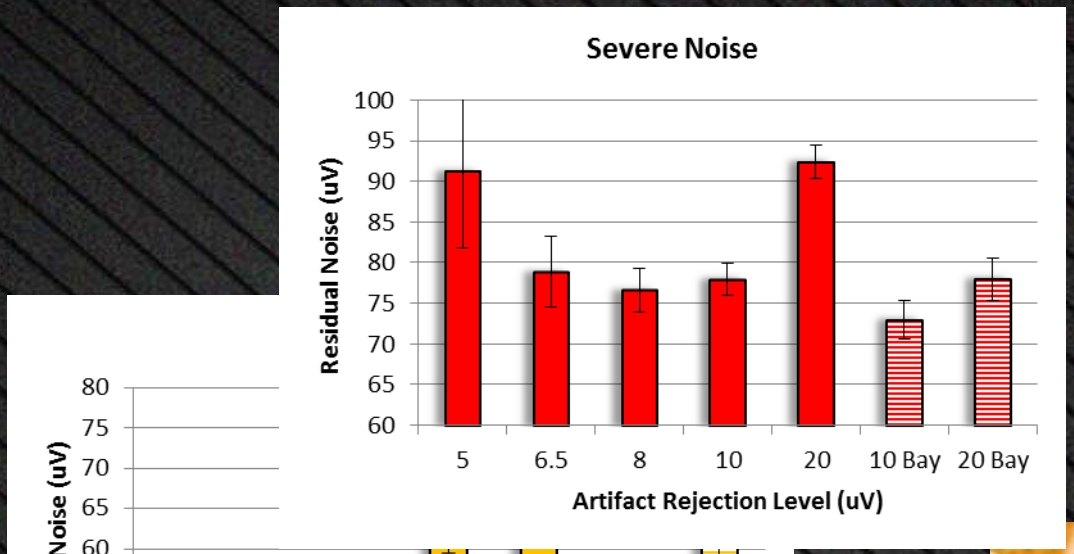
- Test time was fixed (3000 sweeps @ 49.1/s = 61s)
- The most efficient rejection level will give the lowest residual noise in that time
- Residual noise is computed by the ABR system (Interacoustics Eclipse)
- But not all systems measure residual noise….

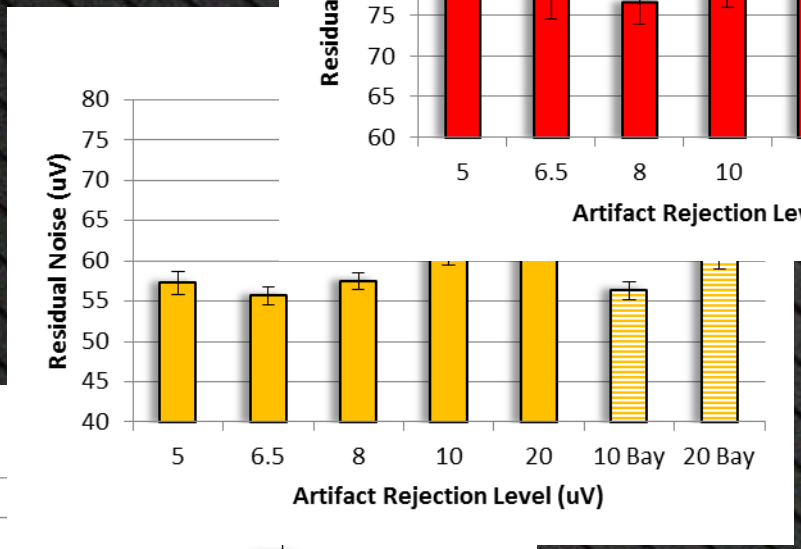
Noise & Rejection

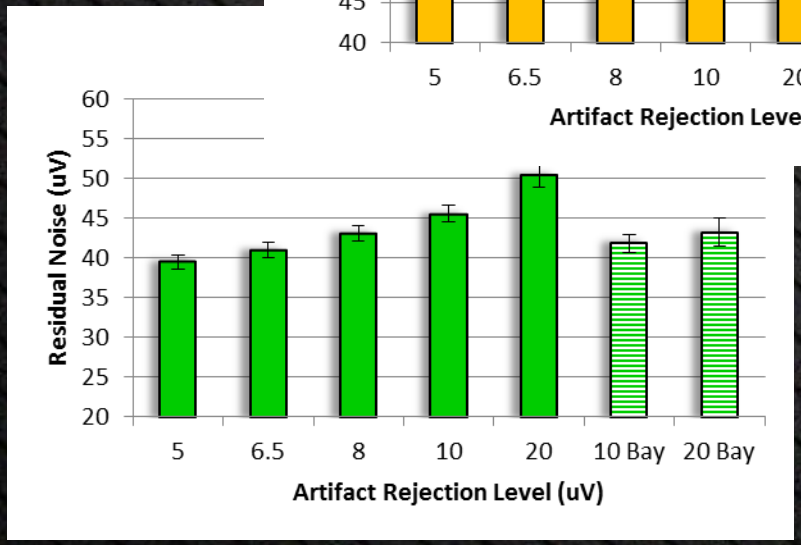
Can use Rejection % as an index of noise



Results - 3 waveform noise categories





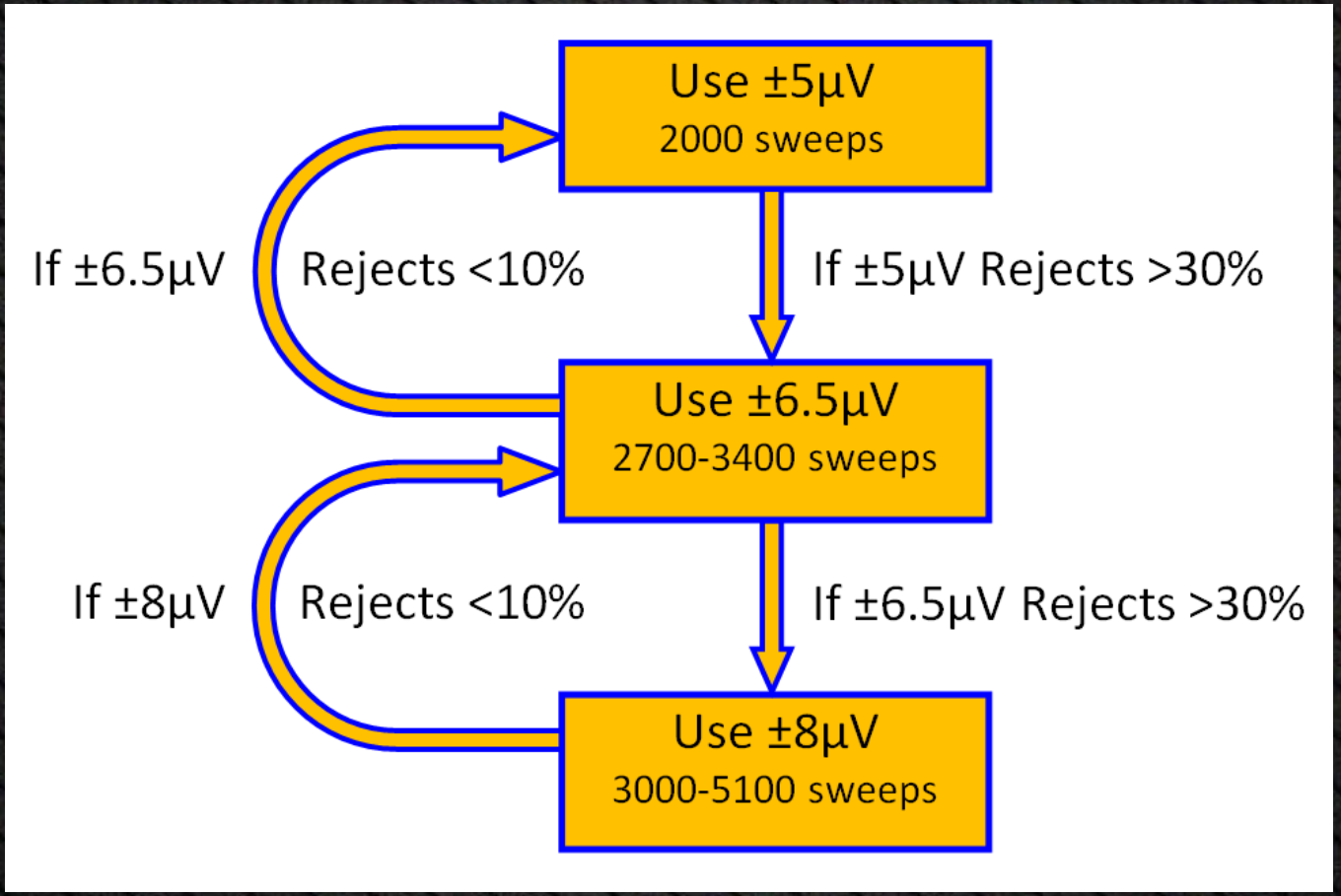


Conclusions of analysis

- In good (low noise) conditions ± 5µV is best
  - Around 2000 sweeps should be adequate
- In moderate noise conditions ± 5µV, 6.5µV & 10µV (with Bayesian) are joint best
  - But to preserve the SNR at 6.5µV ~3000 sweeps are needed
- In severe noise conditions ± 10µV Bayesian is best
- If Bayesian averaging not available, use 6.5µV or 8µV
  - But be prepared to do up to 5000 sweeps at AR= 8µV
- Bayesian averaging helps but is not perfect

A Strategy for testers?

Inspection of the data suggests:



Summary of AR study

- Artefact rejection level affects test efficiency
- The optimum level depends on the extent of noise
- Testers should use a strategy which reflects this
- Use Bayesian averaging if available

Reference: Lightfoot, G., Stevens, J. The effects of artefact rejection and Bayesian weighted averaging on the efficiency of recording the newborn ABR. Ear & Hearing 2014; 35(2): 213-220.

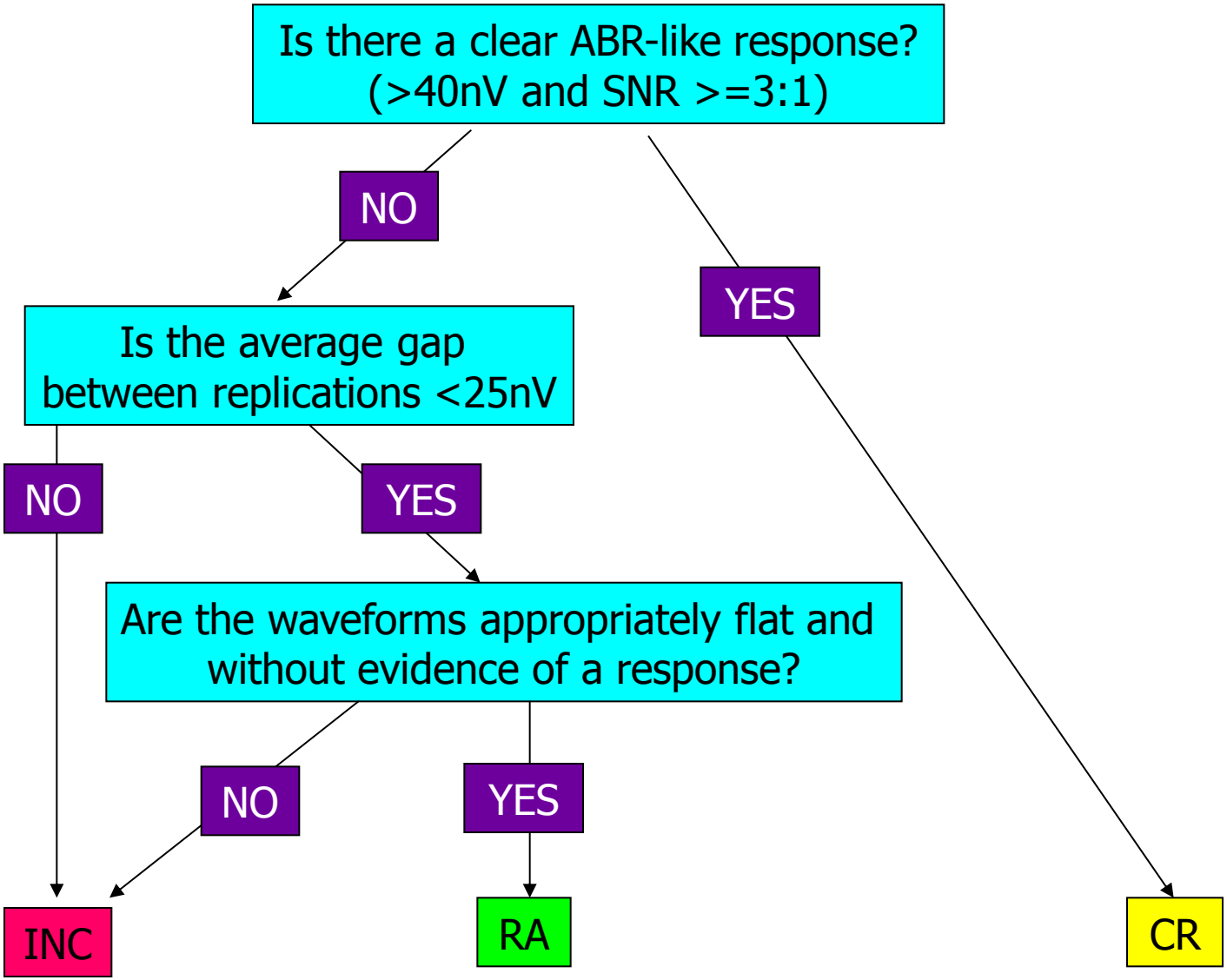
Errors of interpretation

- “Old school” approach:
  a response is either there or it is not
- NHSP approach:
  not 2 but 3 possible outcomes
  - Response is present, with a high degree of certainty (NHSP terminology “Clear Response”, CR)
  - Response is absent, with a high degree of certainty (NHSP terminology “Response Absent”, RA)
  - Recording conditions too poor to tell (NHSP terminology “Inconclusive”, Inc)
- Inconclusive levels *cannot* contribute to the definition of threshold

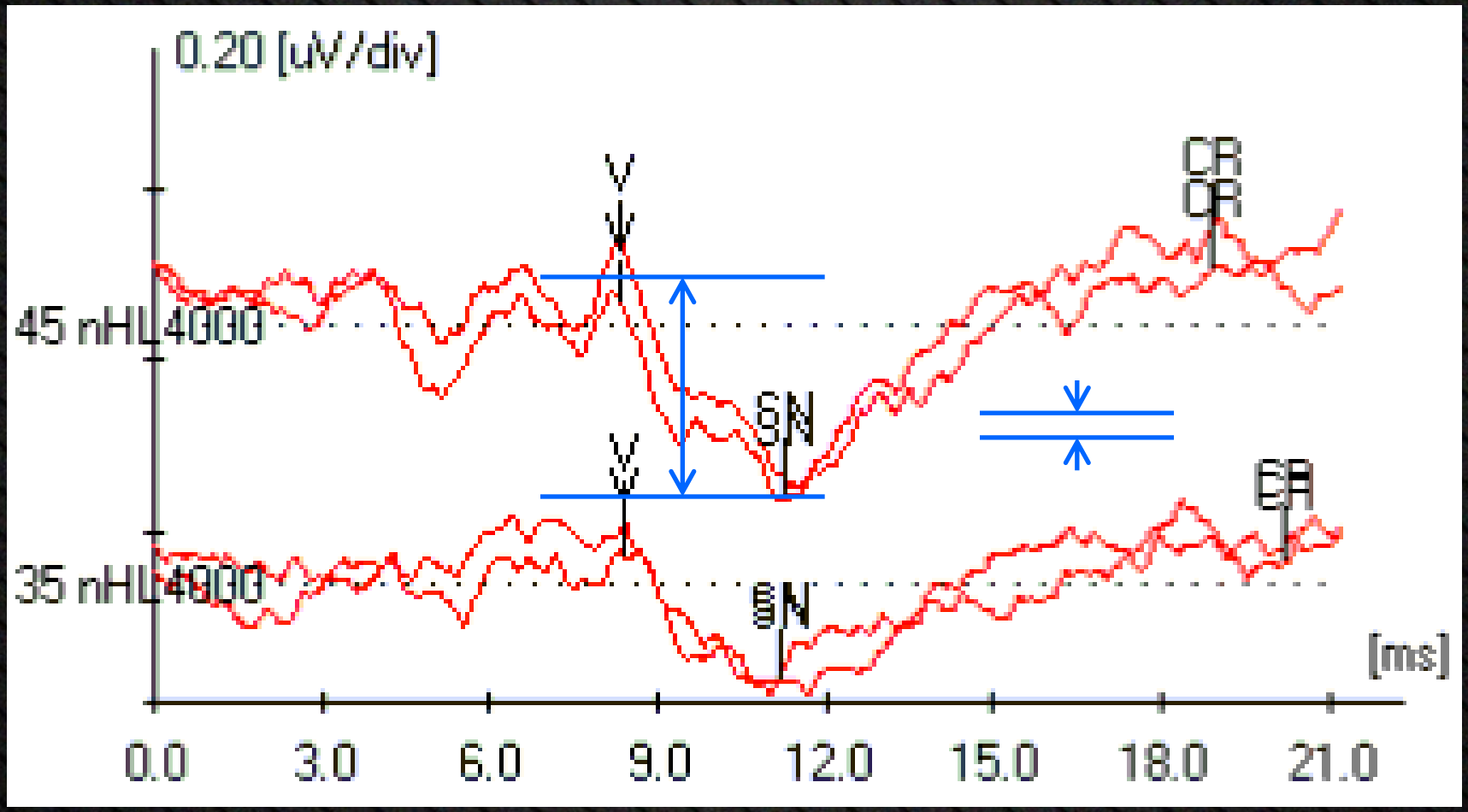
Categorising waveforms: Clear Response - CR

- For a response to be deemed to be present there must be:
  - a high degree of correlation between the replications
  - a characteristic waveform of at least 40nV in size
- The size of the response - judged from top (wave V or wave III) to bottom (SN_{10})- should be at least *3 times the amplitude of the background noise level*
- The noise level can be estimated from <u>average</u> gap between the traces across the recording window
- This criterion ensures a high degree of confidence (about 98%) in the presence of an ABR response

Rating responses at each level: (2013 guidance)



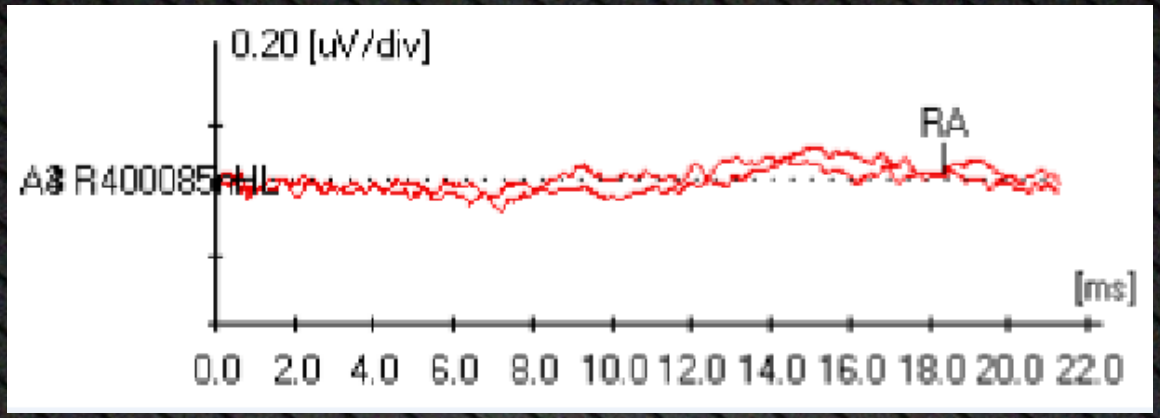
Example – CR

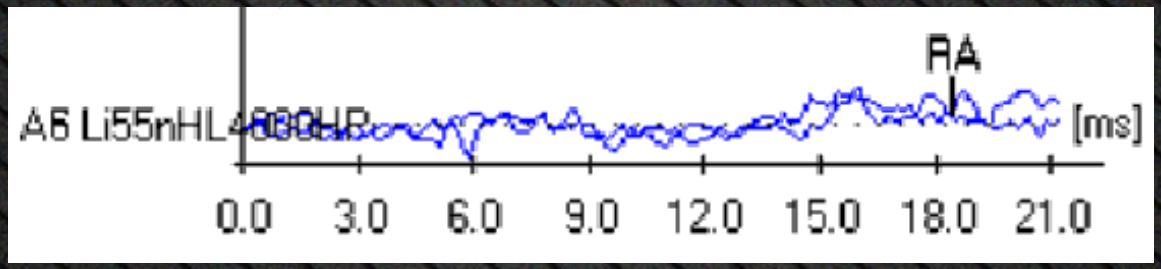

0.20 [uV/div]
V
CR
CR
45 nHL4000
SN
V
CR
35 nHL4000
SN
[ms]
0.0
3.0
6.0
9.0
12.0
15.0
18.0
21.0

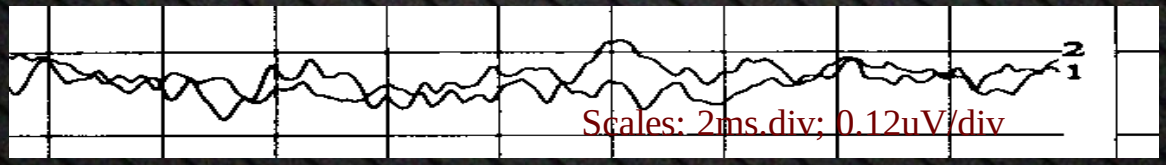
Categorising waveforms
Response Absent - RA

- Superimpose waveforms
- Assess noise as the average gap between replicates over whole window (but ignore any region of stimulus artefact)
- Average gap must be no more than 25nV (0.025µV)
- Tip: the average gap is usually about 1/3 of the maximum gap
- The waveforms must be ‘appropriately flat’ with no evidence of a vestigial response
- This gives a high degree of confidence we are genuinely below threshold

Examples – RA


0.20 [uV/div]
RA
[ms]
0.0 2.0 4.0 6.0 8.0 10.0 12.0 14.0 16.0 18.0 20.0 22.0

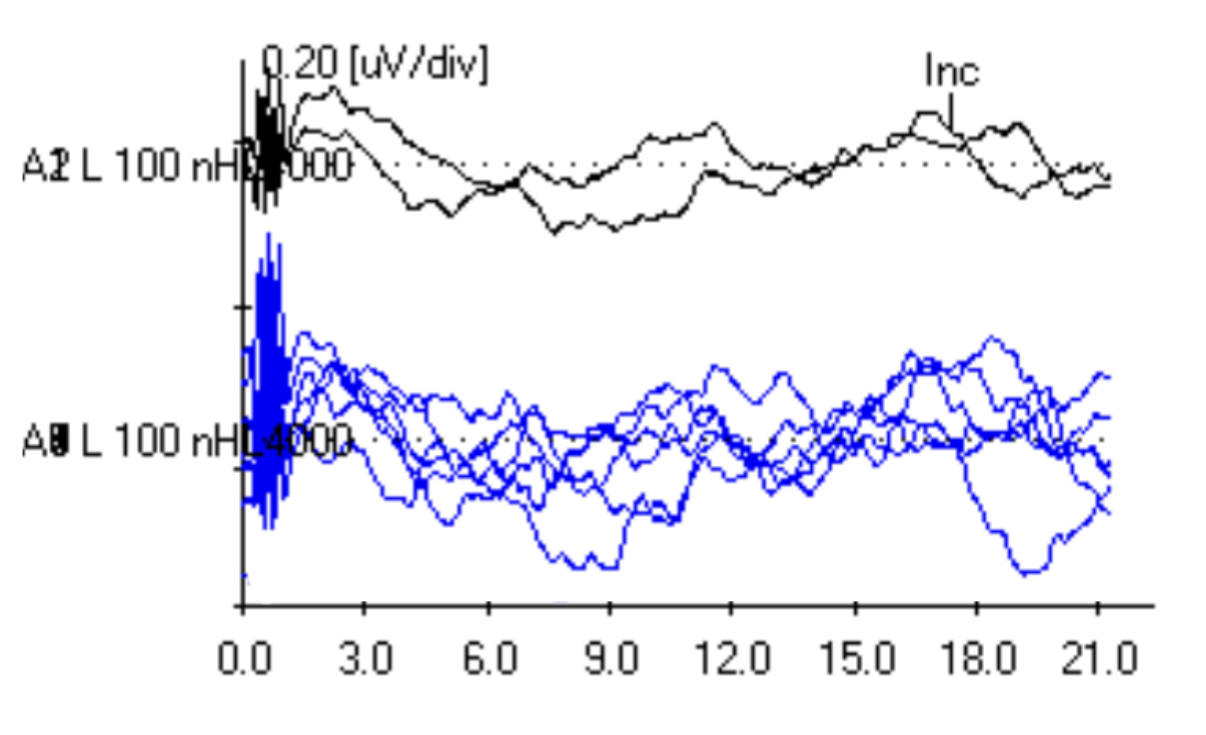

RA
[ms]
0.0 3.0 6.0 9.0 12.0 15.0 18.0 21.0


2
1
Scales: 2ms.div; 0.12uV/div

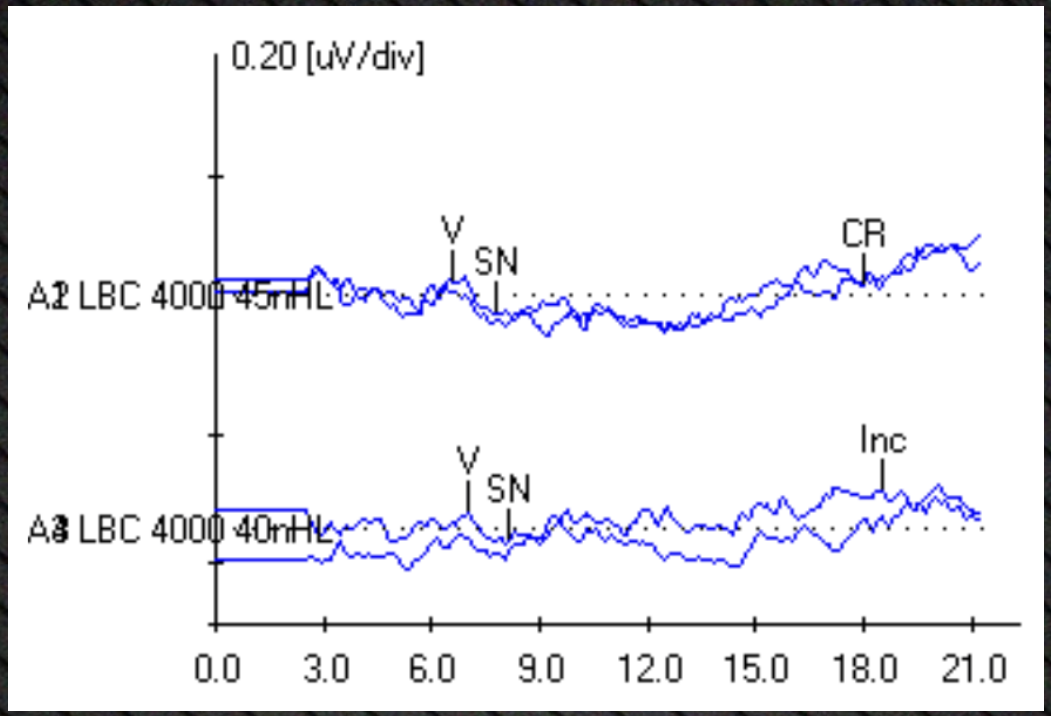
Categorising waveforms
Inconclusive - Inc

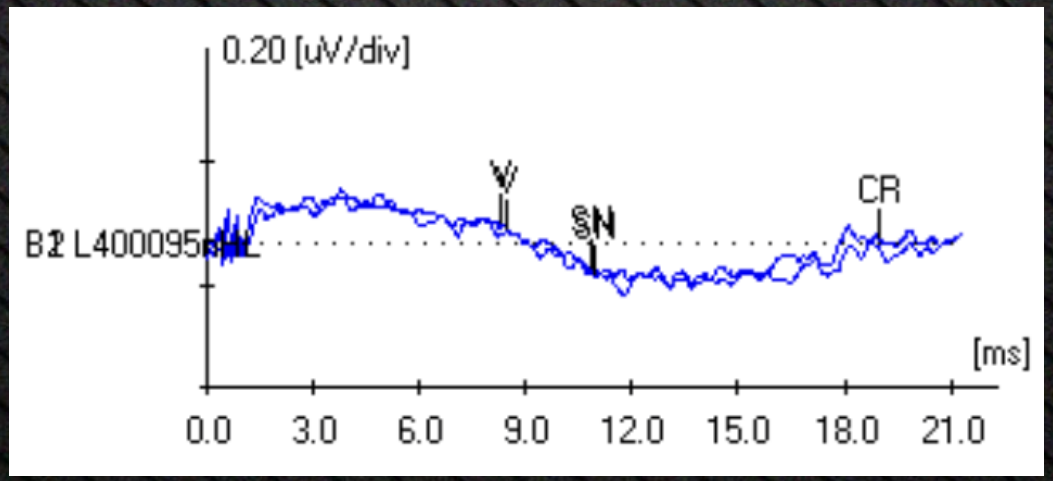
- All other waveforms are “inconclusive”

  - the replications will have S/N < 3:1 or have no obvious response yet have noise greater than the criterion value



Examples - Inc





Consequences of labelling CR when no response is present

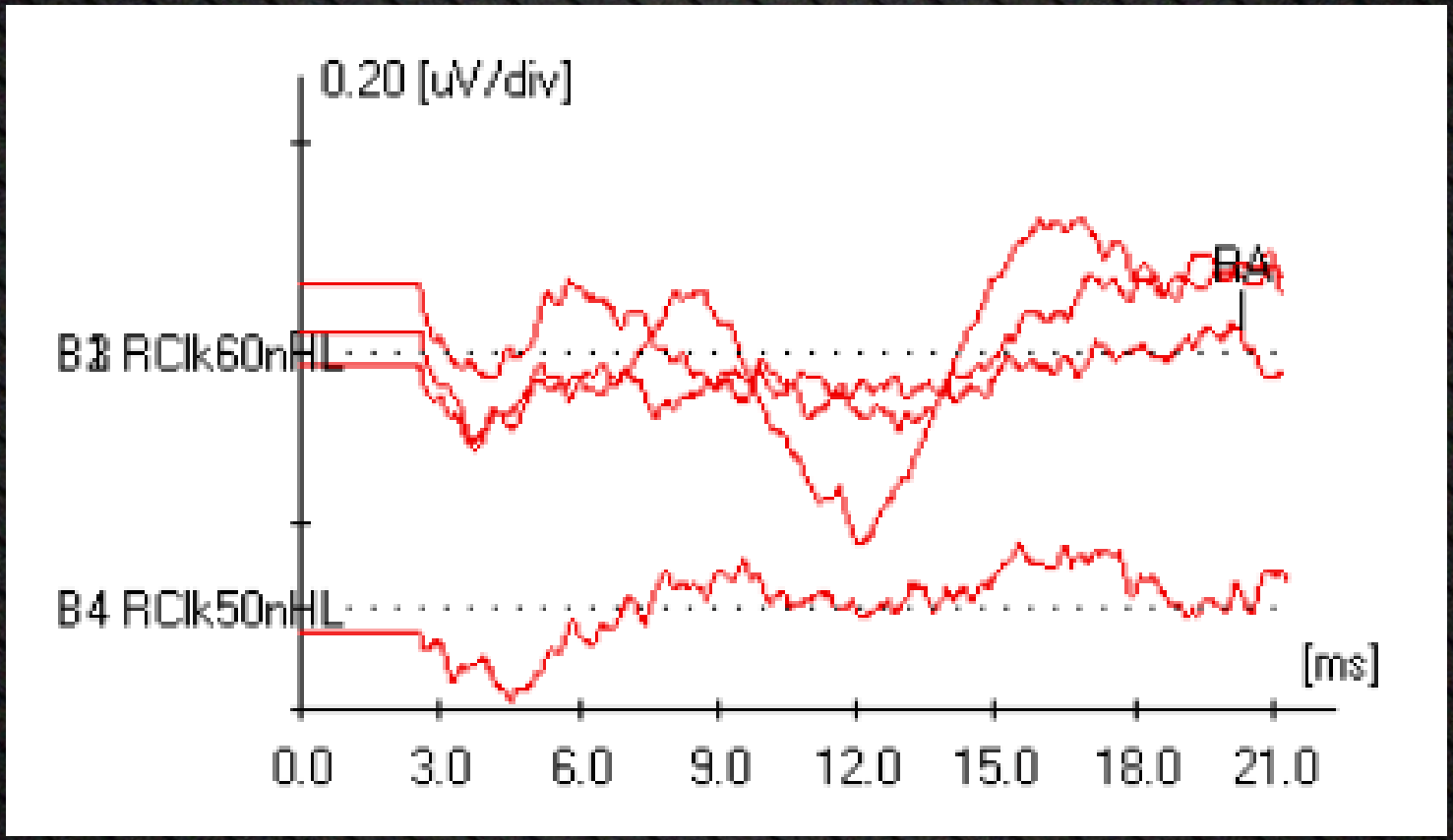
- Noise is mistaken for a response
- Discharge child with hearing loss
  - Worst-case: label a profoundly deaf child as normal
  - Child is lost to follow-up
  - Eventually discovered, too late
  - Legal case could ensue
- Underestimate hearing threshold for a PCHI: under-amplification

Consequences of labelling RA when response is present

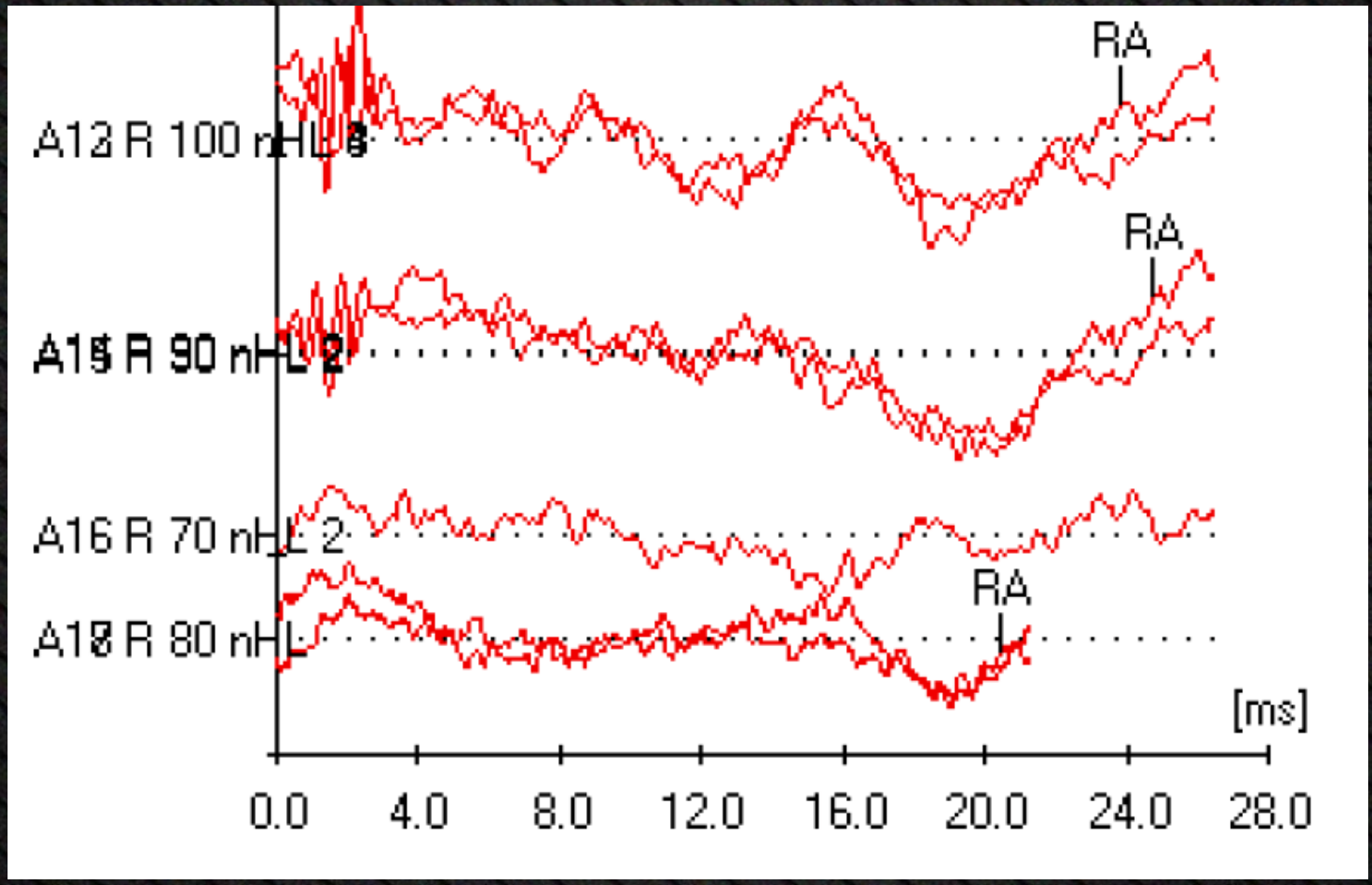
- Response is buried in noise
- Identify normal-hearing child as having hearing loss
- Overestimate hearing threshold for a PCHI: over-amplification
- Worst case: aid a child with normal hearing

Time for some howlers, interesting and difficult cases, all revealed in the QA process….

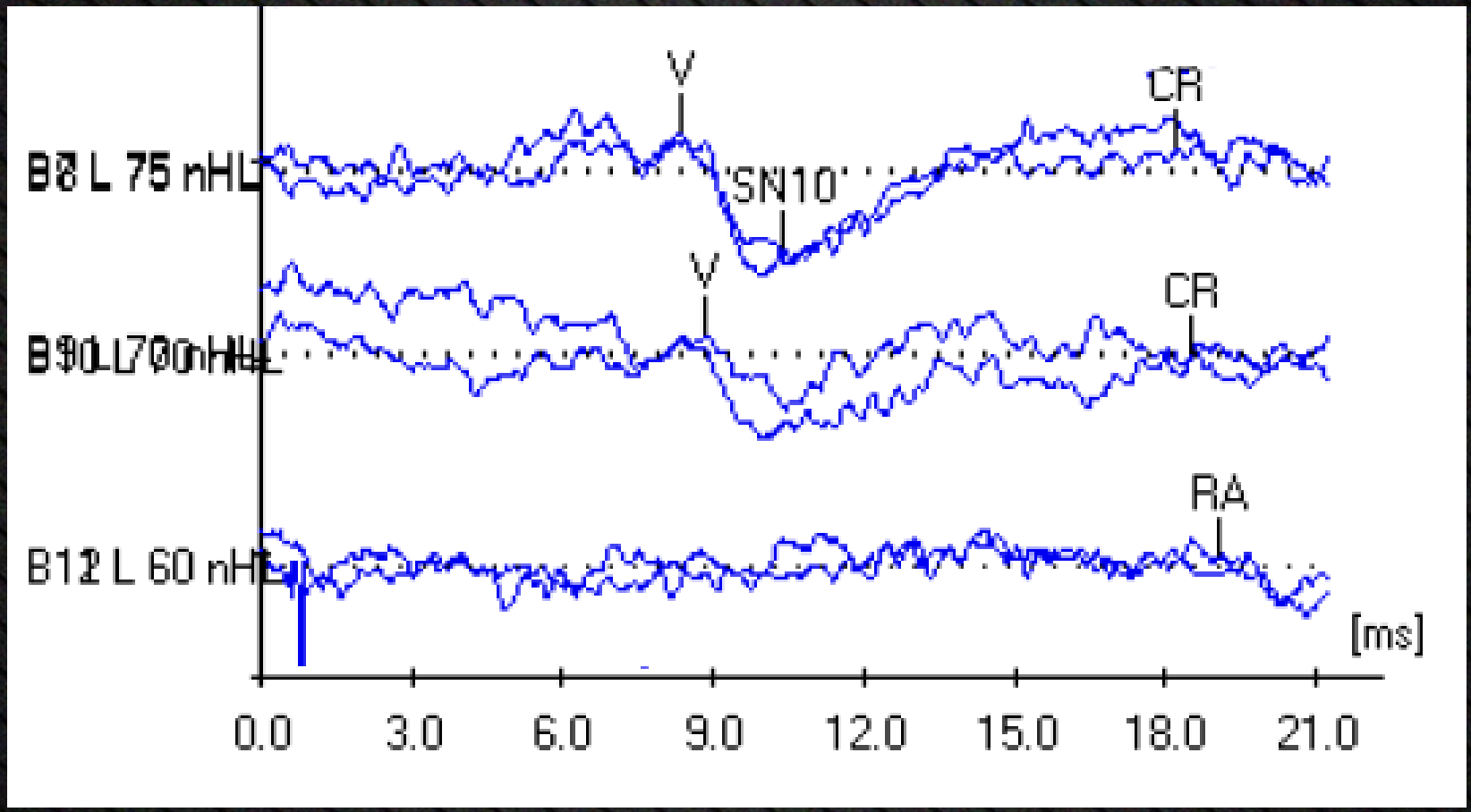
ABR Example 1 (Click) “>60dB”



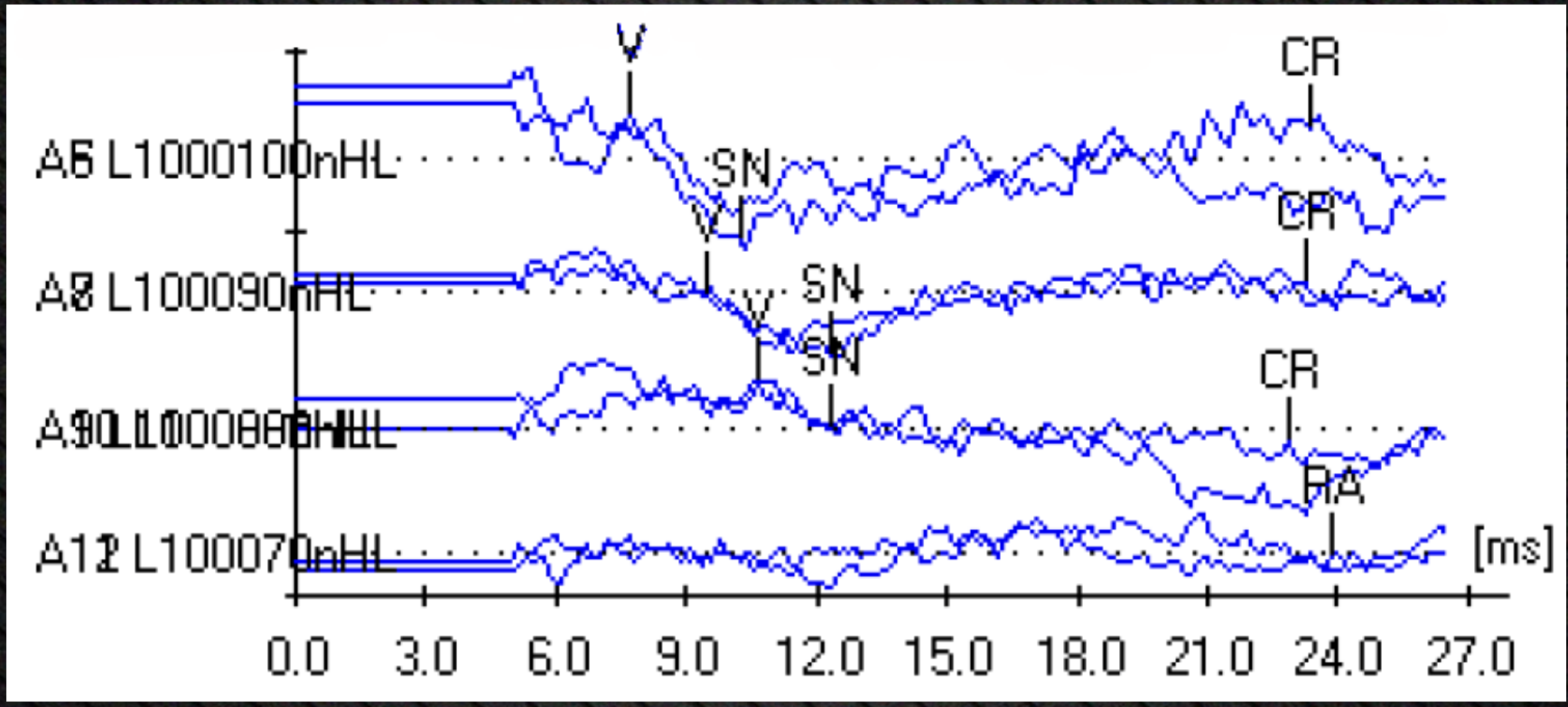
ABR Example 2 (4k & 1k)



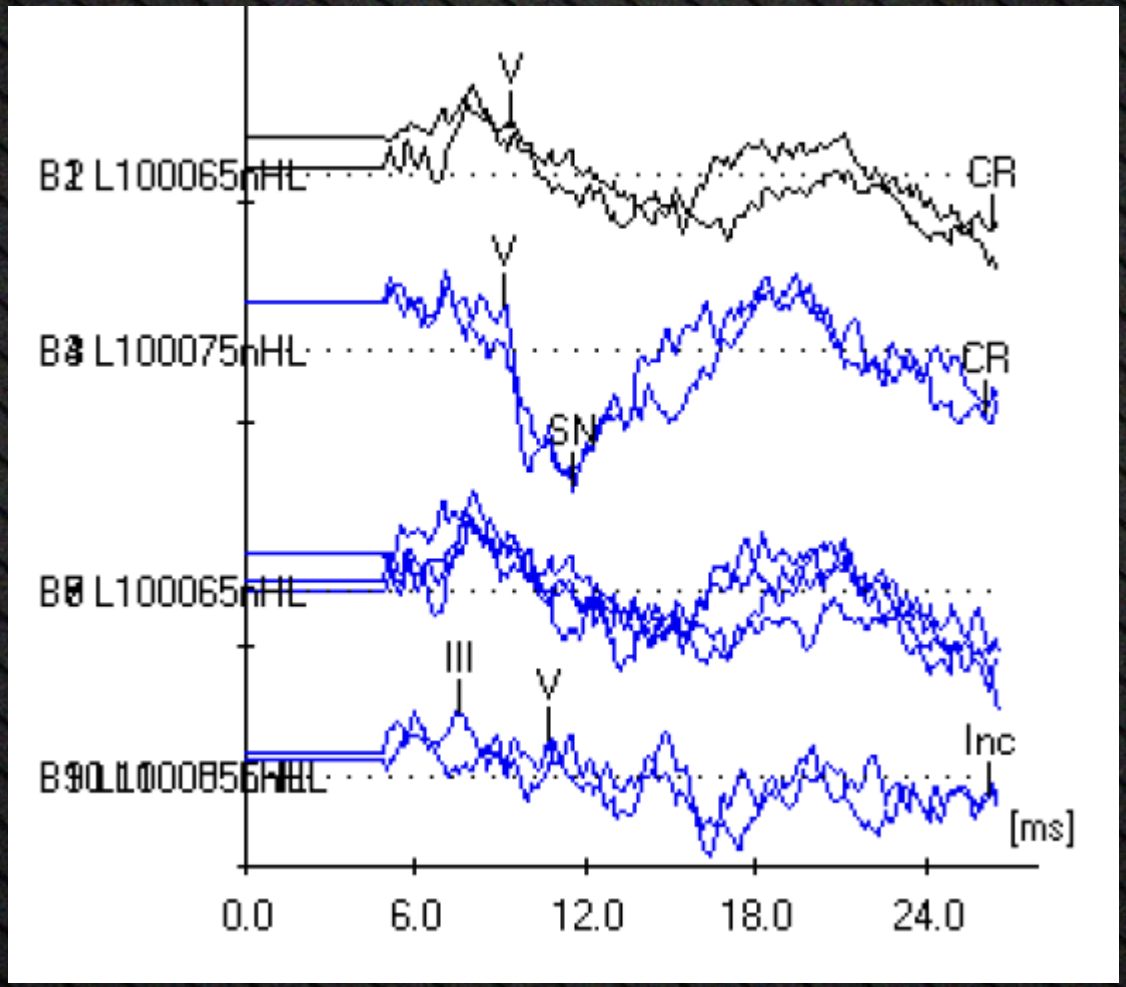
ABR Example 3 (4k) “=70”



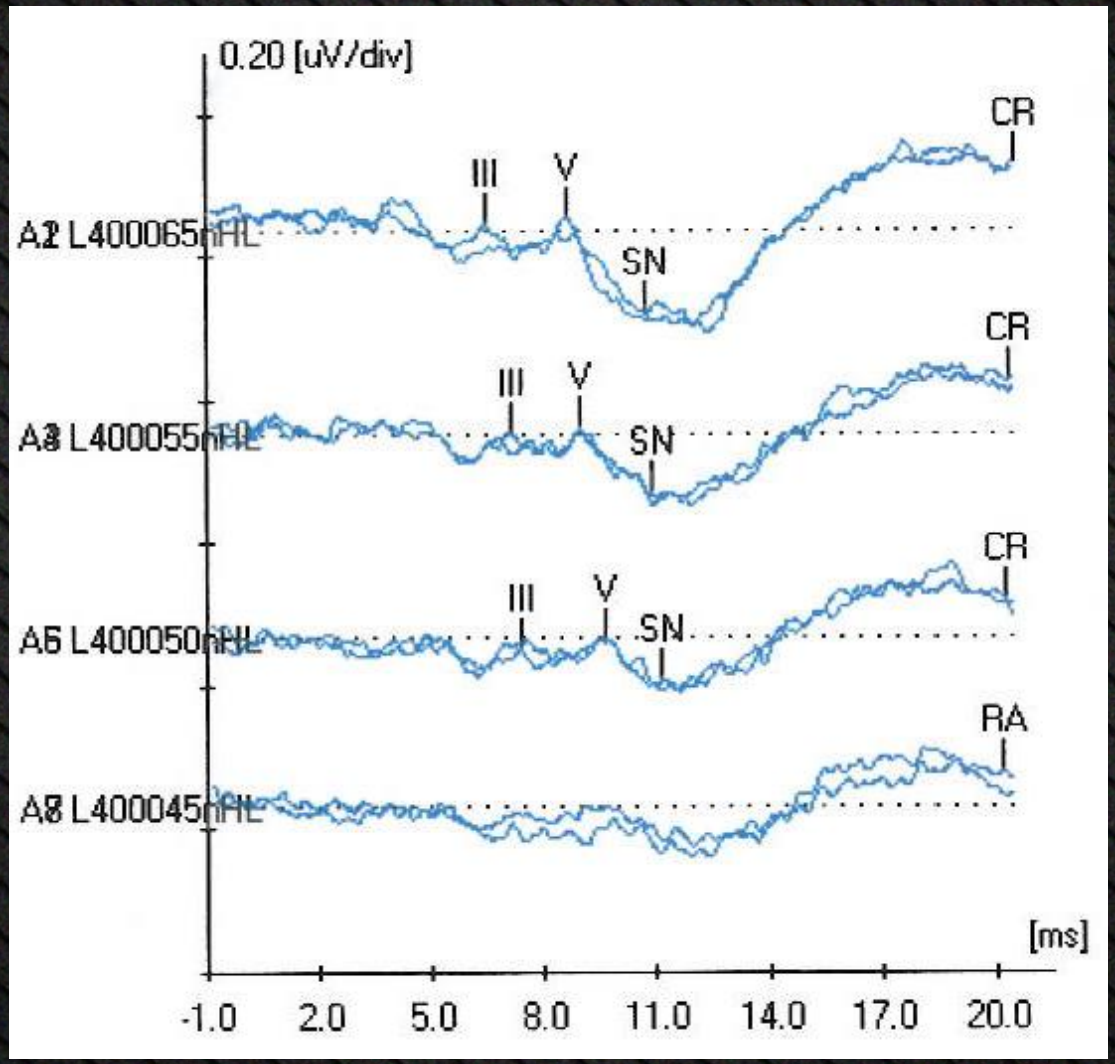
ABR Example 4 (1kHz) “=80dB”



ABR Example 5 (1kHz) “≤65dB”



ABR Example 6 (4kHz) “=50dB”



Errors of test strategy

- Starting with the “wrong” ear
  - NHSP policy to start a unilateral referral with the passed ear (no test is perfect; it’s important to verify that at least one ear is satisfactory – language acquisition will be dominated by the status of the better ear)
- Poor stimulus level selection
  - reduces efficiency & scope of what is achieved in each test session
  - NHSP guidance: start at discharge level +10dB
- Failing to apply masking when needed
- Failing to perform BC testing (can’t rely on tymps)
- Failing to perform CM testing when needed

Errors of reporting

- Reporting e.g. =65dBnHL instead of ≤65dBnHL when no “RA” is obtained
  - In theory the ear could be normal
- Not conveying limitations in test precision
  - Test conditions may have compromised results – this must be included in the clinical report
- Incorrect dBnHL to dBeHL correction
  - Corrections depend on frequency, age and transducer
- Transposing ears
  - Getting Rt & Lt ears mixed up; could lead to inappropriate amplification

Errors of case management

- Unnecessary delay in testing
  - ABR more likely to be problematic after 12 weeks
- Failure to follow-up when appropriate
- Failure to re-test possible ANSD cases
  - Many resolve (presumed delayed neurological maturation)
- Premature amplification or implantation in ANSD

Would we change our guidance?

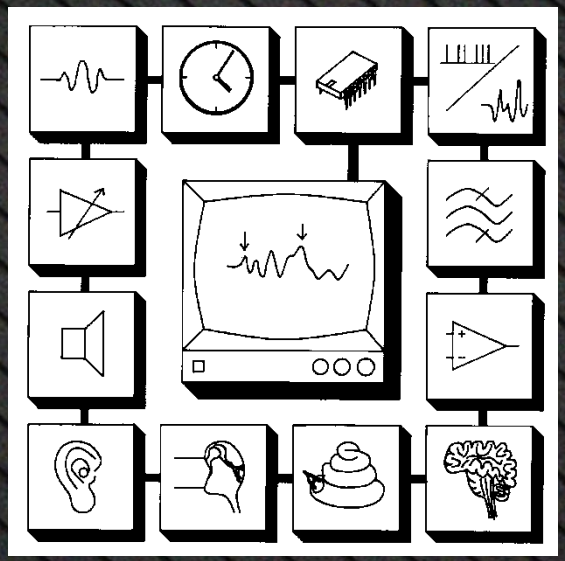
- If all ABR systems offered objective measures there would be no need to replicate
- Instead of fixed number of sweeps, tester would average until appropriate to stop – in the *prevailing conditions*
- Fsp or SNR would help identify response presence & therefore when to stop averaging
- Residual noise would help identify when noise is low enough to conclude response absence
- Both CR & RA would still require tester judgement
  - CR: response morphology/size; RA no evidence of a response

Improving ABR standards

- Available options to facilitate improvement:
  - Require ABR testers to undergo certification (driving test)
  - Free “refresher” courses for all ABR testers
  - On-site visits to identify issues and initiate re-training
  - Close monitoring / mentoring of worst performing sites
  - Suspend service of sites resistant to change
  - Encourage the development of regional peer review groups, with national moderation & support
  - Introduce remote “tele-audiometry” ABR or on-line expert
  - Continue QA audits to monitor quality

- The talk tomorrow will reveal what NHSP did - and what they should have done but didn’t!

Many thanks for your attention!



www.eratraining.co.uk
www.abrpeerreview.co.uk