



Regarding sources of irreducible variation in transformed up-down staircase experiments

Andrew Christian

Structural Acoustics Branch
NASA Langley Research Center

177th Meeting of the Acoustical Society of America
1pPP11, May 13th 2019, Louisville, KY

The Technical Memorandum



- This work stems from a battery of tests undertaken at the NASA Langley Research Center to measure low-frequency auditory filters.
- Presented by Rafaelof previously:
<https://doi.org/10.1121/1.5014759>
- The full report is available on the NASA Technical Reports Server:
<https://ntrs.nasa.gov/search.jsp?R=20190001050>

NASA/TM-2019-220120



Characterization of Low Frequency Auditory Filters

Menachem Rafaelof
National Institute of Aerospace, Hampton, Virginia

Andrew W. Christian and Kevin P. Shepherd, Stephen A. Rizzi
Langley Research Center, Hampton, Virginia

James H. Stephenson
US Army Aviation and Missile Research,
Development and Engineering Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

January 2019

The Technical Memorandum



- The goal was to determine *population averaged* low frequency auditory filter shapes.
 - The protocol included testing a large number of people, each rather minimally.
- The results were presented by Rafaelof previously:

<https://doi.org/10.1121/1.5014759>

NASA/TM-2019-220120



Characterization of Low Frequency Auditory Filters

Menachem Rafaelof
National Institute of Aerospace, Hampton, Virginia

Andrew W. Christian and Kevin P. Shepherd, Stephen A. Rizzi
Langley Research Center, Hampton, Virginia

James H. Stephenson
US Army Aviation and Missile Research,
Development and Engineering Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

January 2019

The Technical Memorandum



- A considerable effort was made to design the test beforehand – quantification of sources of variance in order to determine number of subjects, number of test points, etc.
- This is captured in a (substantial) appendix (B) in the document.

NASA/TM-2019-220120



Characterization of Low Frequency Auditory Filters

Menachem Rafaelof
National Institute of Aerospace, Hampton, Virginia

Andrew W. Christian and Kevin P. Shepherd, Stephen A. Rizzi
Langley Research Center, Hampton, Virginia

James H. Stephenson
US Army Aviation and Missile Research,
Development and Engineering Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

January 2019

Overview



- This presentation will primarily cover things that are discussed in that appendix.
 - This includes several concepts that became central to the test design that do not seem to be reflected in psychophysical/-acoustic literature.
- There will be some signal detection theory thrown in at the end...

NASA/TM-2019-220120



Characterization of Low Frequency Auditory Filters

Menachem Rafaelof
National Institute of Aerospace, Hampton, Virginia

Andrew W. Christian and Kevin P. Shepherd, Stephen A. Rizzi
Langley Research Center, Hampton, Virginia

James H. Stephenson
US Army Aviation and Missile Research,
Development and Engineering Center, Hampton, Virginia

National Aeronautics and
Space Administration

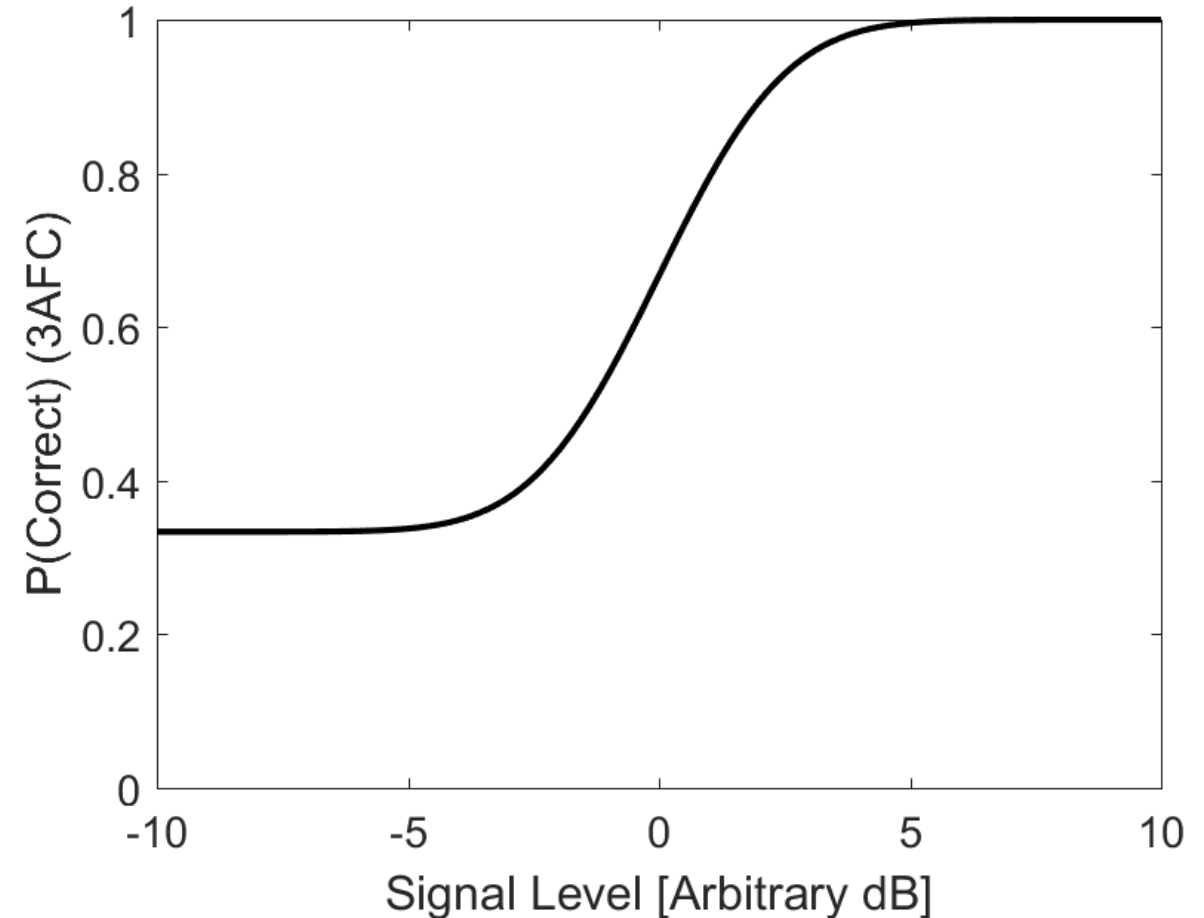
Langley Research Center
Hampton, Virginia 23681-2199

January 2019

Whence Irreducible Variance

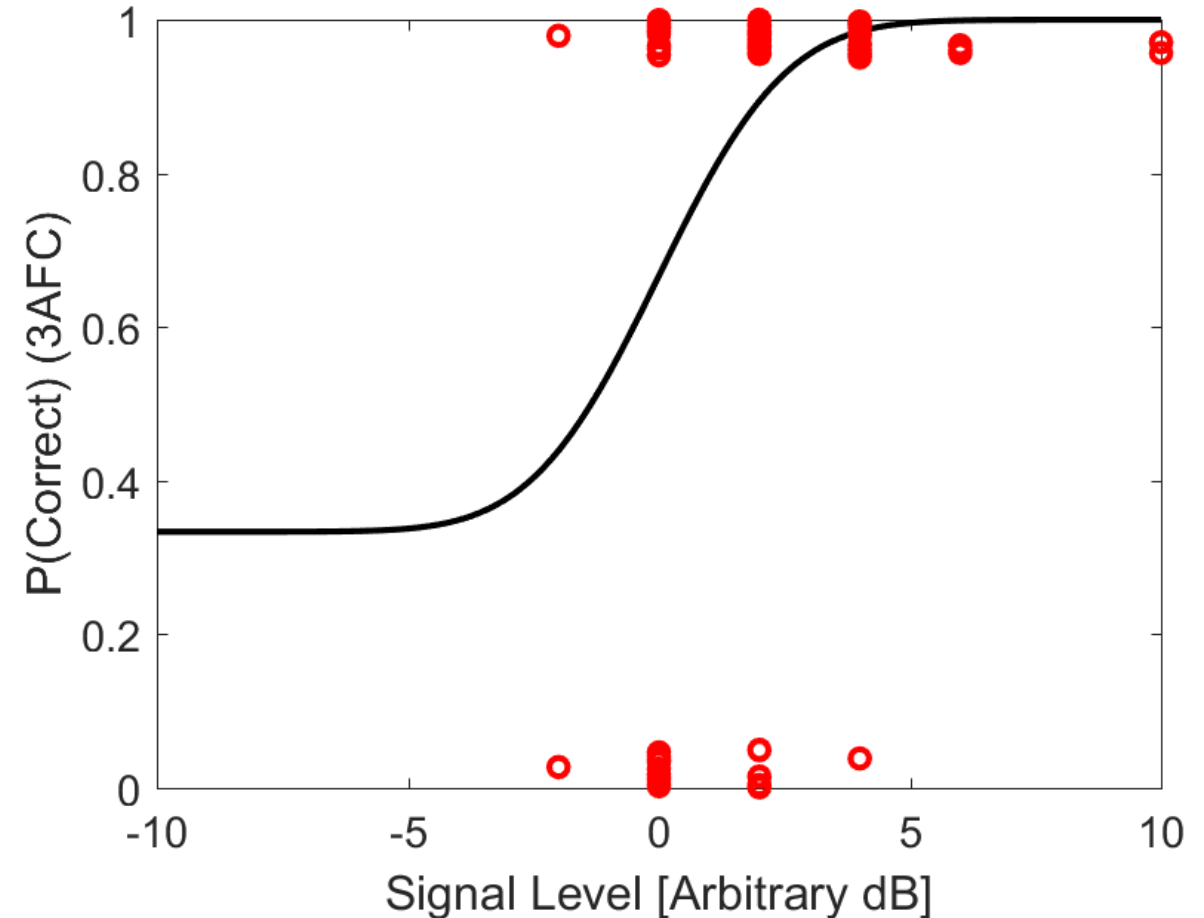


- First, consider the psychometric function.
- This describes the probability a subject has correctly answering an n-alternative forced-choice (nAFC) question (trial).



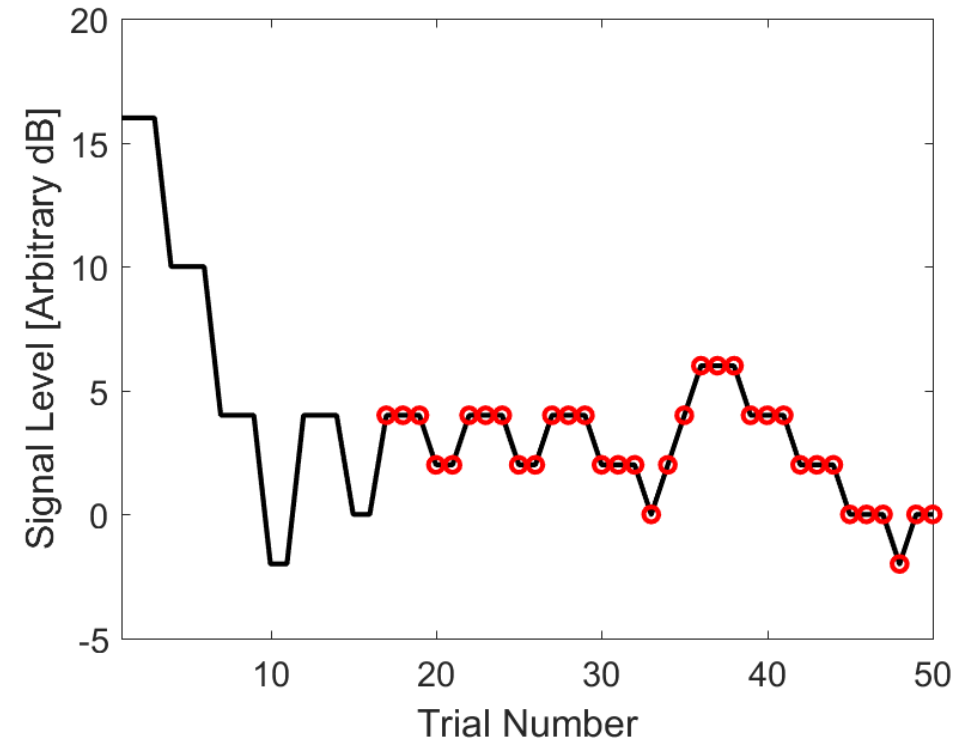
Whence Irreducible Variance

- Consider the ‘perfect’ subject – one whose psychometric function doesn’t change over time.
 - I.e., with the effects of learning, attention, etc.
- Using nAFC questions, the responses of that subject would still be *discretized* or *quantized* into correct/incorrect binary data.
 - They could get the trial incorrect in places that they were very likely to have heard the signal, and vice-versa.



Whence Irreducible Variance

- In this study, transformed up-down staircases were used to figure out at which signal levels to place the trials.
- Generally, the algorithms of the staircase are x-down, 1-up (xD1U), indicating how many trials a subject needs to get right or wrong to move the signal level.

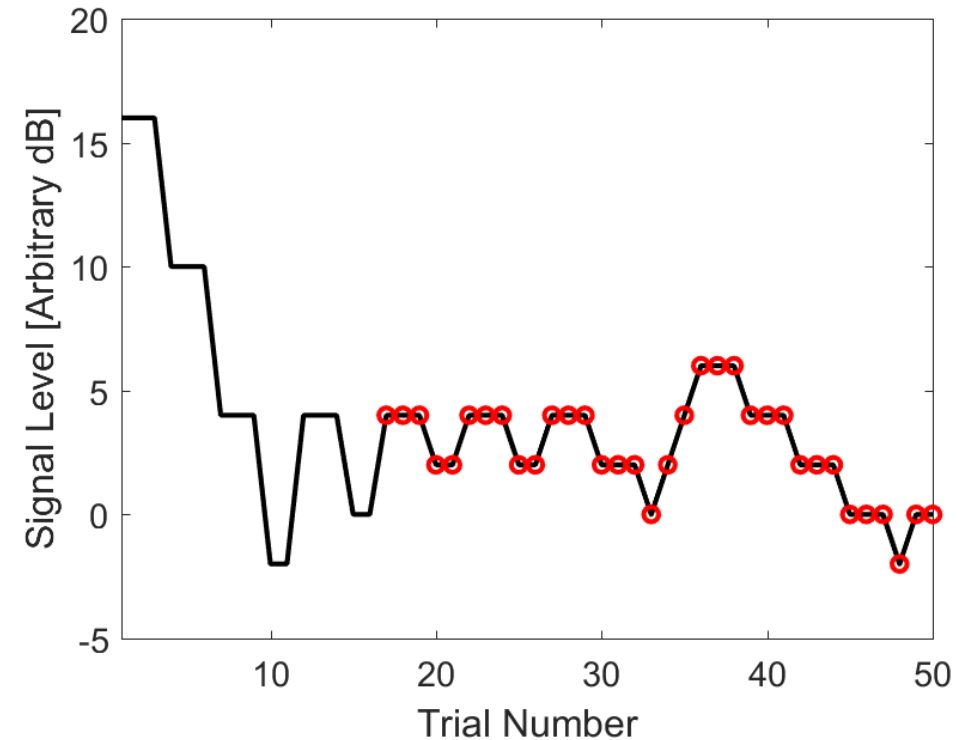


Whence Irreducible Variance

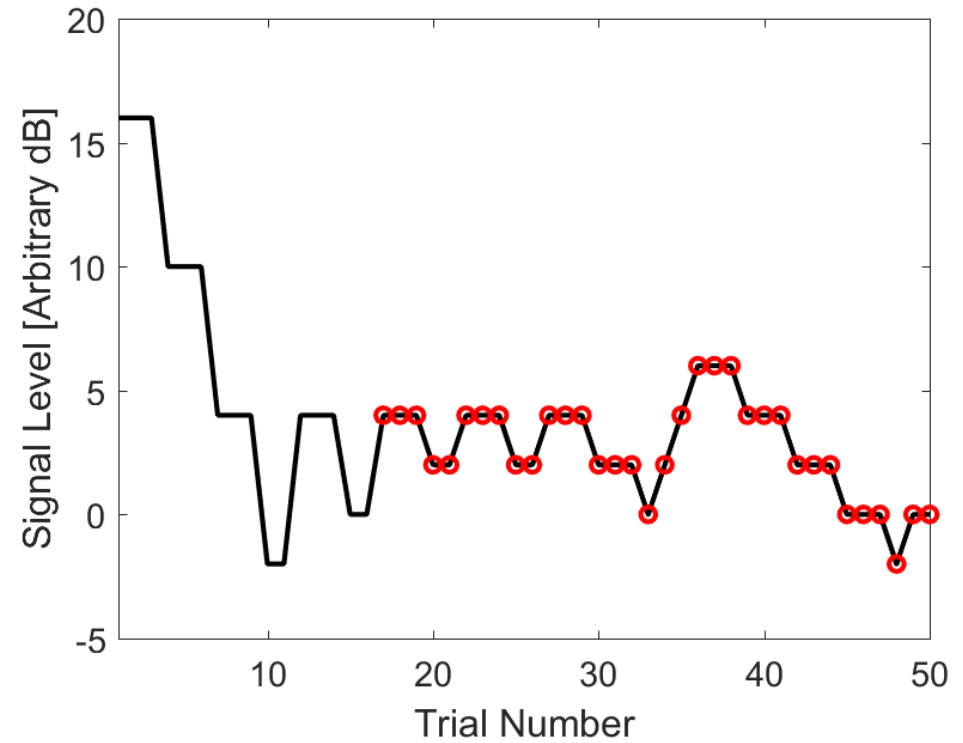
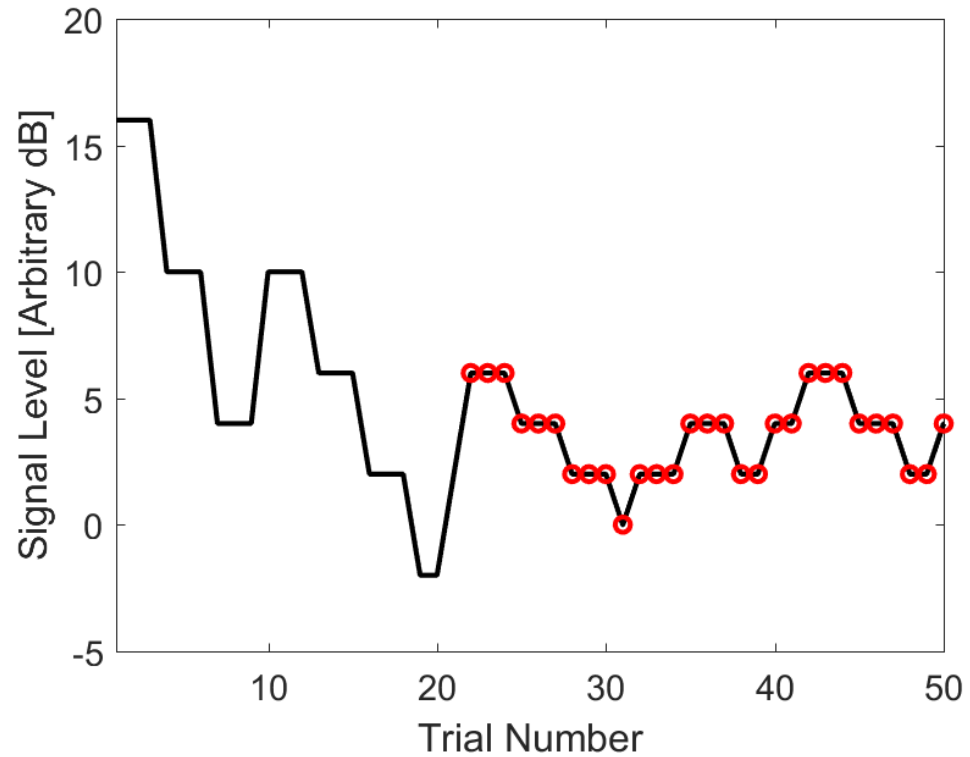
- Due to the discretized nature of the responses, even for the perfect subject, the evolution of the staircase will be randomly different every time it is run.

→ These differences imply a distribution of the point estimates generated from the data.

The standard deviation of this distribution is σ_{SC} .



Whence Irreducible Variance





Designing a Test

- What we want to know, in order to design a test, is what is the optimum combination of $n(\text{AFC})$ and $x(\text{D1U})$ in terms of σ_{SC} ?
 - What will produce the fastest test, in terms of minimum σ_{SC} -per-time?
 - What is a sufficient number of trials for that staircase?
 - What level of variation can we expect?
- This winds up being a very complicated question to answer, as there are a number of factors that all interact.
 - We can simulate many staircases for the perfect subject.

Determining σ_{SC}

- The results shown are for simulations of 100,000 staircases for:

$n(\text{AFC}) =$

— 2

— 3

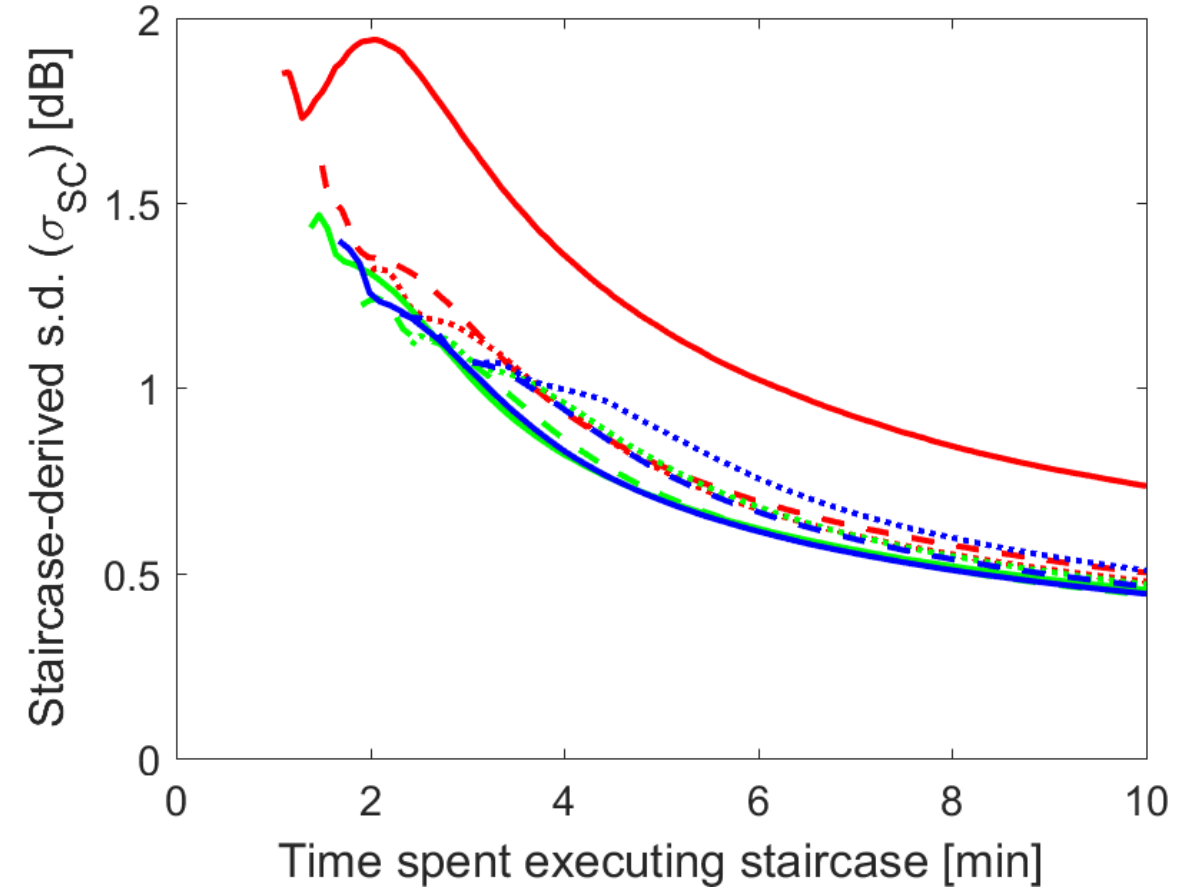
— 4

$x(\text{D1U}) =$

— 2

- - - 3

..... 4

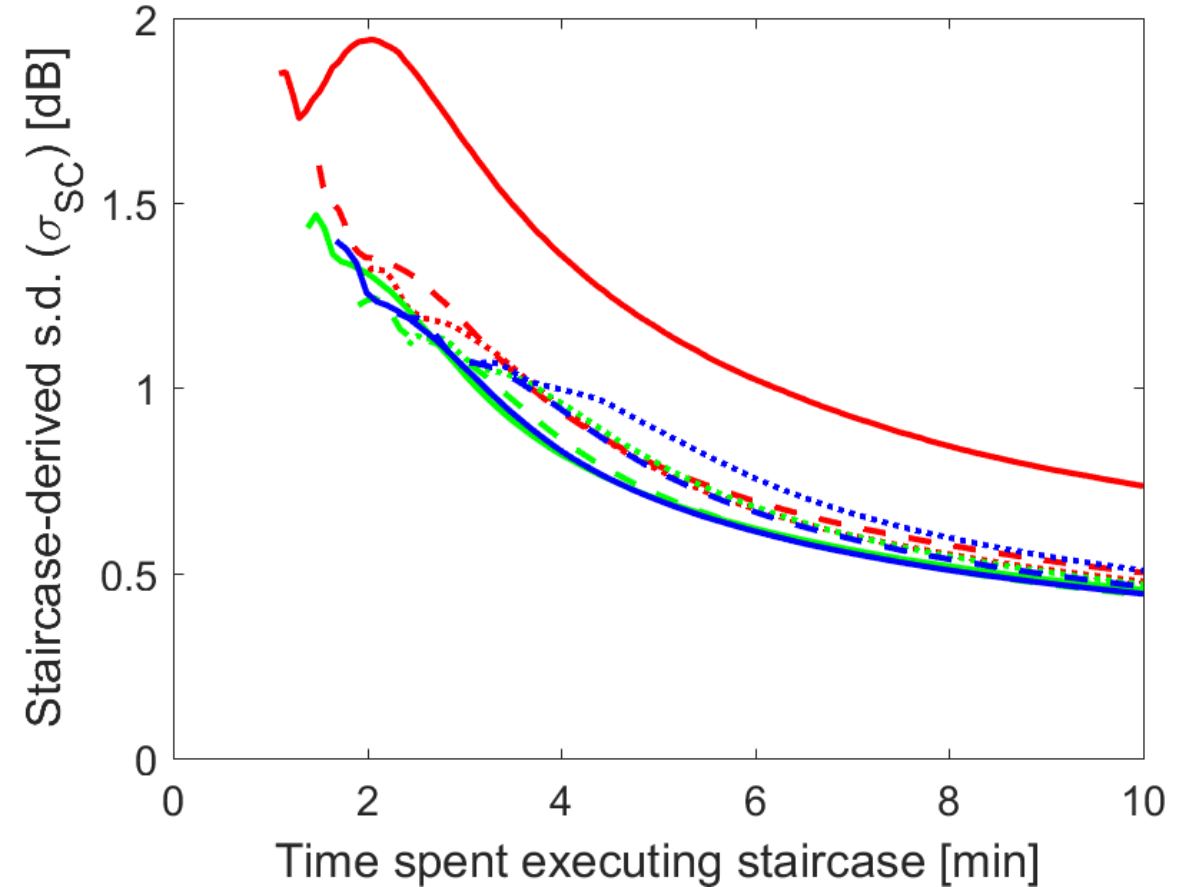


Determining σ_{SC}

Observations:

- 2AFC procedures are all sub-optimal
- The best combinations are nearly indistinguishable.
- After about 5 minutes of testing, for *any* combination, σ_{SC} begins to level off.

→ 3AFC/3D1U was settled on with consideration of other practical issues.



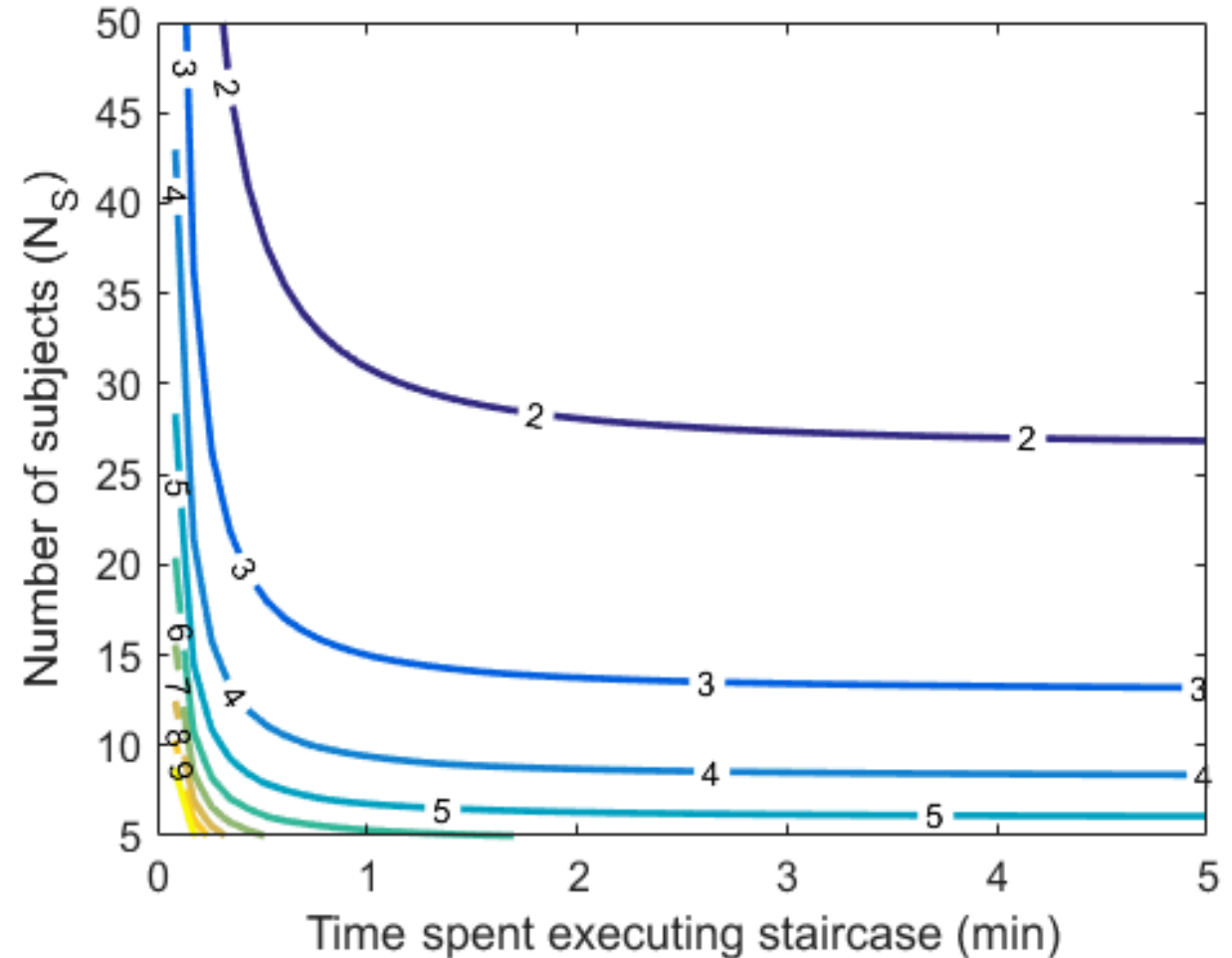
Designing a Test

- Using an estimate of between-subject variation ($\sigma_{BS} = 5$ dB, which is reasonable at low frequencies), one can determine how much sway σ_{SC} will have over the overall results.

$$\sigma_{TOTAL} = \sqrt{\sigma_{BS}^2 + \sigma_{SC}^2}$$

- The contours shown are of the predicted confidence interval size for the overall result.

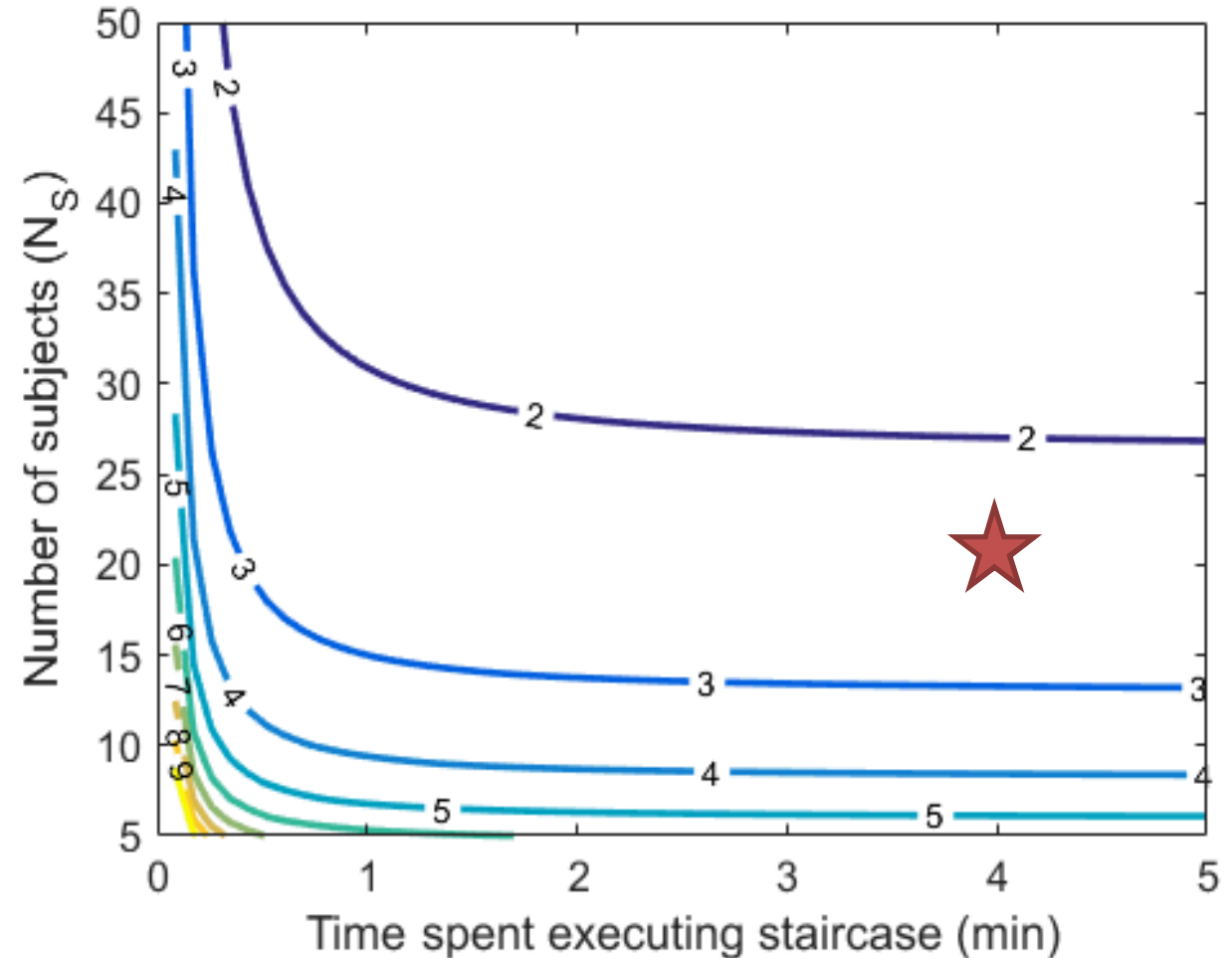
Contours of Predicted Confidence Interval Radius [dB]



Designing a Test

- 50 trial staircases were used – only 4 minutes long.
 - Staircases have to be minimally long to converge (at least 3 or 4 minutes)
 - σ_{SC} is around 1 dB
- 20 subjects completed 12 staircases each to map out a complete auditory filter in a single hour-long visit.
 - Subjects came from the population surrounding NASA Langley and had minimal training.

Contours of Predicted Confidence Interval Radius [dB]





Comparison with Repeated Measures

- 10 of the subjects also completed 3 repeats of a test condition
- The measured value of $\sigma_{WITHIN} = 1.6$ dB
 - There are other significant sources of intra-subject variation (as should be expected), but, on average, about half of this variation is explained by σ_{SC} .
- The measured value of $\sigma_{BETWEEN} = 3.8$ dB, less than the 5 dB that was assumed. The measured σ_{TOTAL} was close to the prediction.



Comparison with Repeated Measures

Interestingly, 5 of the subjects display $\sigma_{WITHIN} < 1$ dB, and the others show $\sigma_{WITHIN} > 2$ dB – there were no subjects between 1 and 2 dB!

- This may indicate that the propensity to produce inconsistent results might be a feature of some subjects and not others.
- Though, it's important to keep in mind that this is a small sample of 10 subjects and 3 repeats.

Conclusions



- The irreducible intra-subject variation of nAFC-based staircase tests was evaluated.
- This allowed the design of a psychoacoustic test that made a non-canonical tradeoff between the number of subjects and the precision to which each subject is tested.
- In the results of the test, on average, half of the intra-subject variability is explained by the irreducible variation.
 - Again, for relatively untrained subjects from the general population.

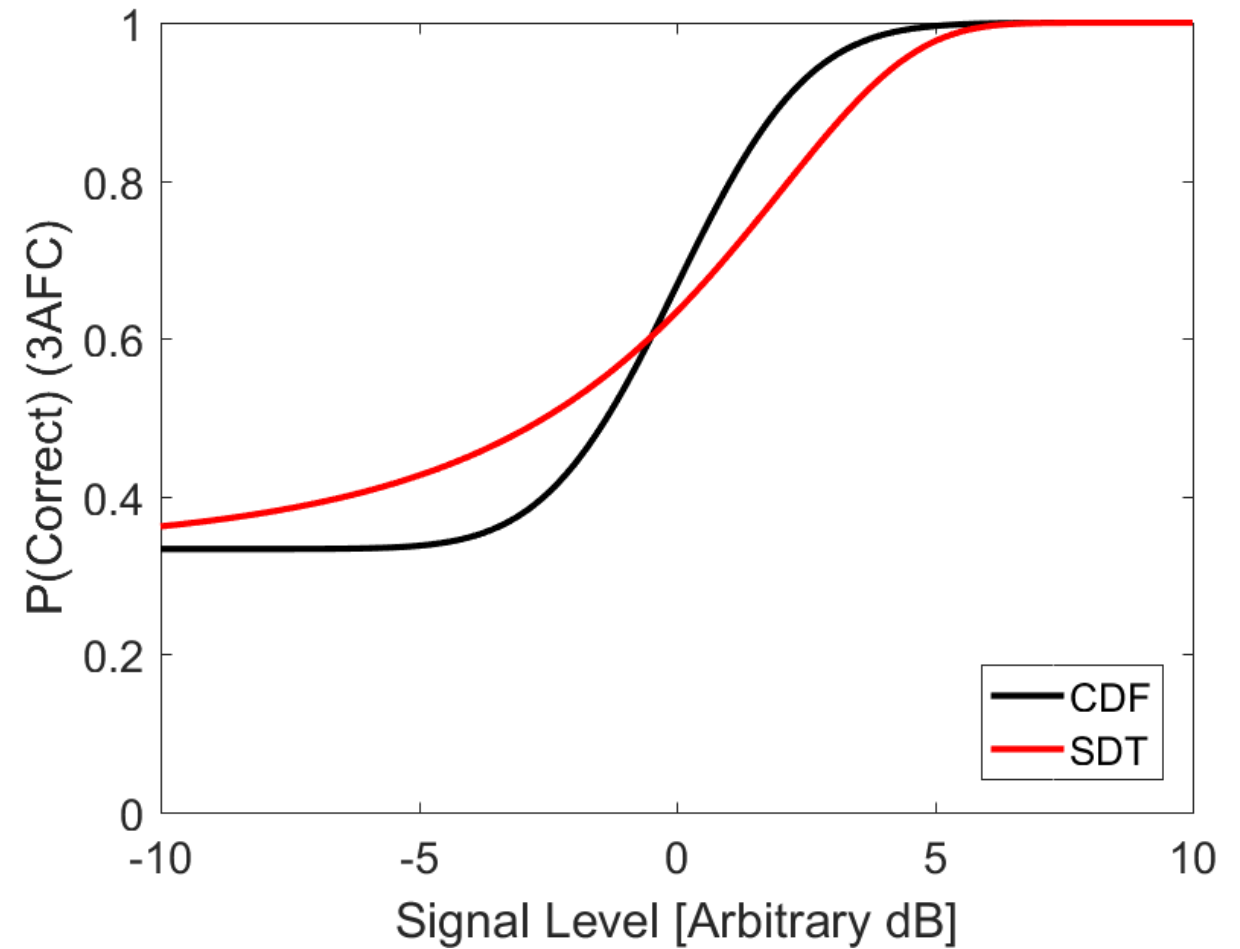
Updated Conclusions



- Some subjects seemed to have a propensity for non-repeatability.
 - The more consistent subjects are likely performing at a level where the only significant source of variability is purely statistical.
- None of this is necessarily at odds with what's in the literature (e.g., that 3AFC/3D1U is the most *time* efficient staircase), but it seems difficult to find this information presented from the point of view of experimental design.

The SDT Psychometric Function

- Compared to the CDF, the predicted tone-in-noise function is considerably shallower.
 - This will lead to more scatter in the result in terms of dB.
- The resultant σ_{SC} is 40% larger.
 - This does not change the selection of 3AFC/3D1U.
 - This does not greatly alter any conclusions.



Acknowledgements



- This work is part of a larger battery of fundamental psychoacoustic tests that have been in development at NASA for the past 4 years. Others who have been crucial to these projects include:
 - Menachem Rafaelof, Kevin Shepherd, Stephen Rizzi, Matt Boucher, James Stephenson and Charlie Smith.
- Publications regarding other facets of this work – involving not only psychoacoustics, but statistics for psychophysics in general (and SDT) – are forthcoming.

Questions

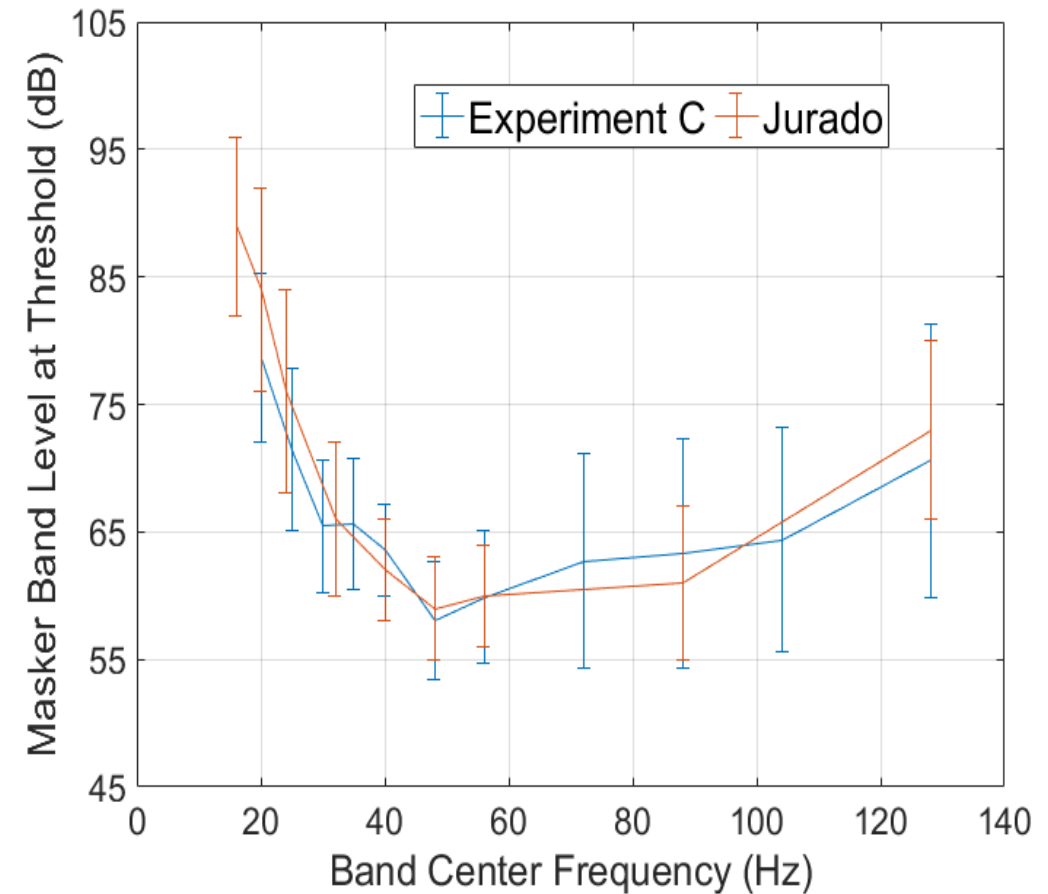


andrew.christian@nasa.gov

<https://ntrs.nasa.gov/search.jsp?R=20190001050>

For the Record

- The ICHIN-I test mapped the lowest auditory filter for 20 subjects (among other measurements).
 - Good agreement with previously published data.
- Why this is amazing:
 - It did not use a purpose-built facility.
 - It used this non-canonical tradeoff of number of subjects and precision-per-subject.
 - The subjects used were minimally trained and taken from the general population.

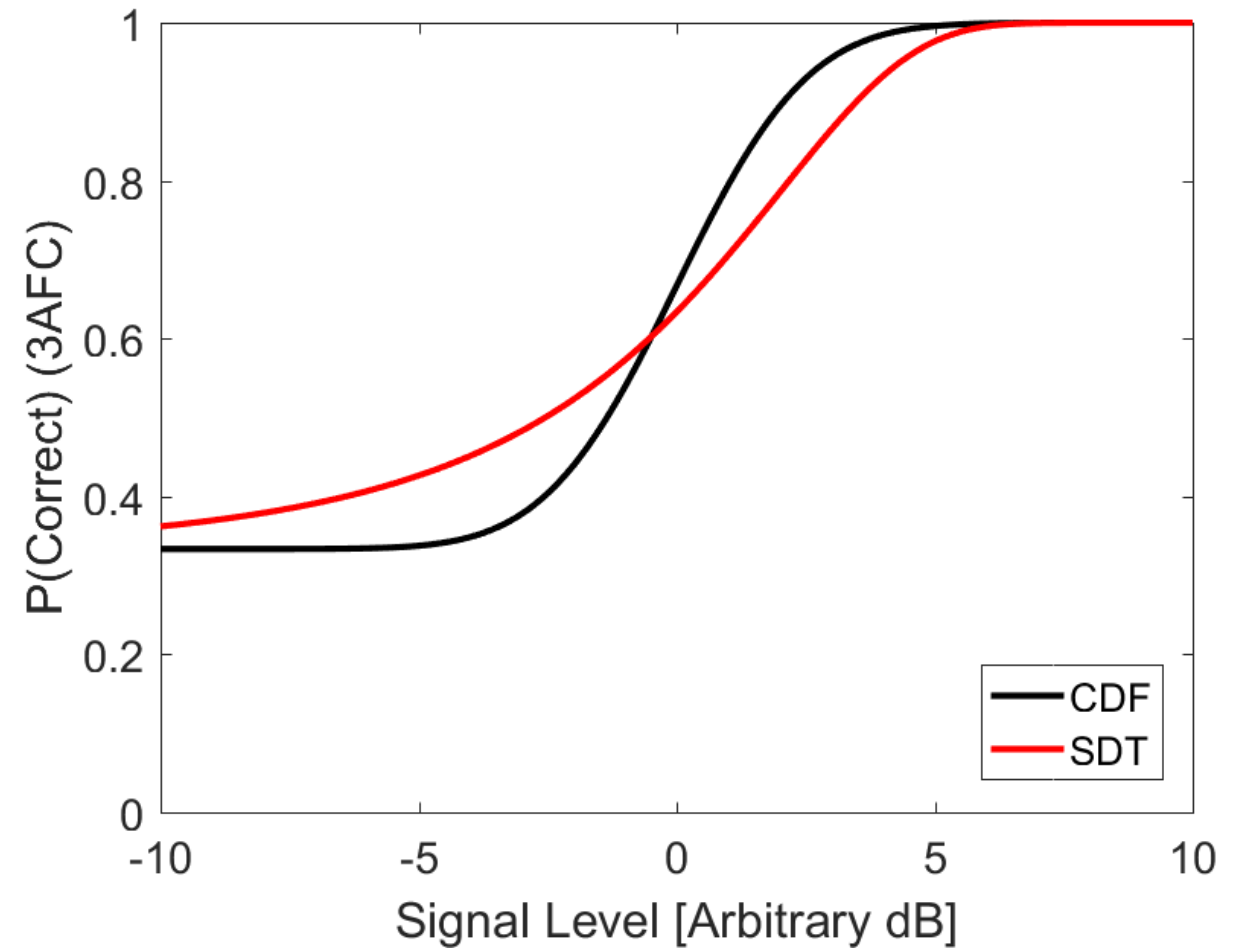


The SDT Psychometric Function

- Everything above used a scaled normal-CDF as a psychometric function .
- Signal detection theory (SDT) gives us a better approximation:
 - High-frequency/Low-signal/Noise-in-noise

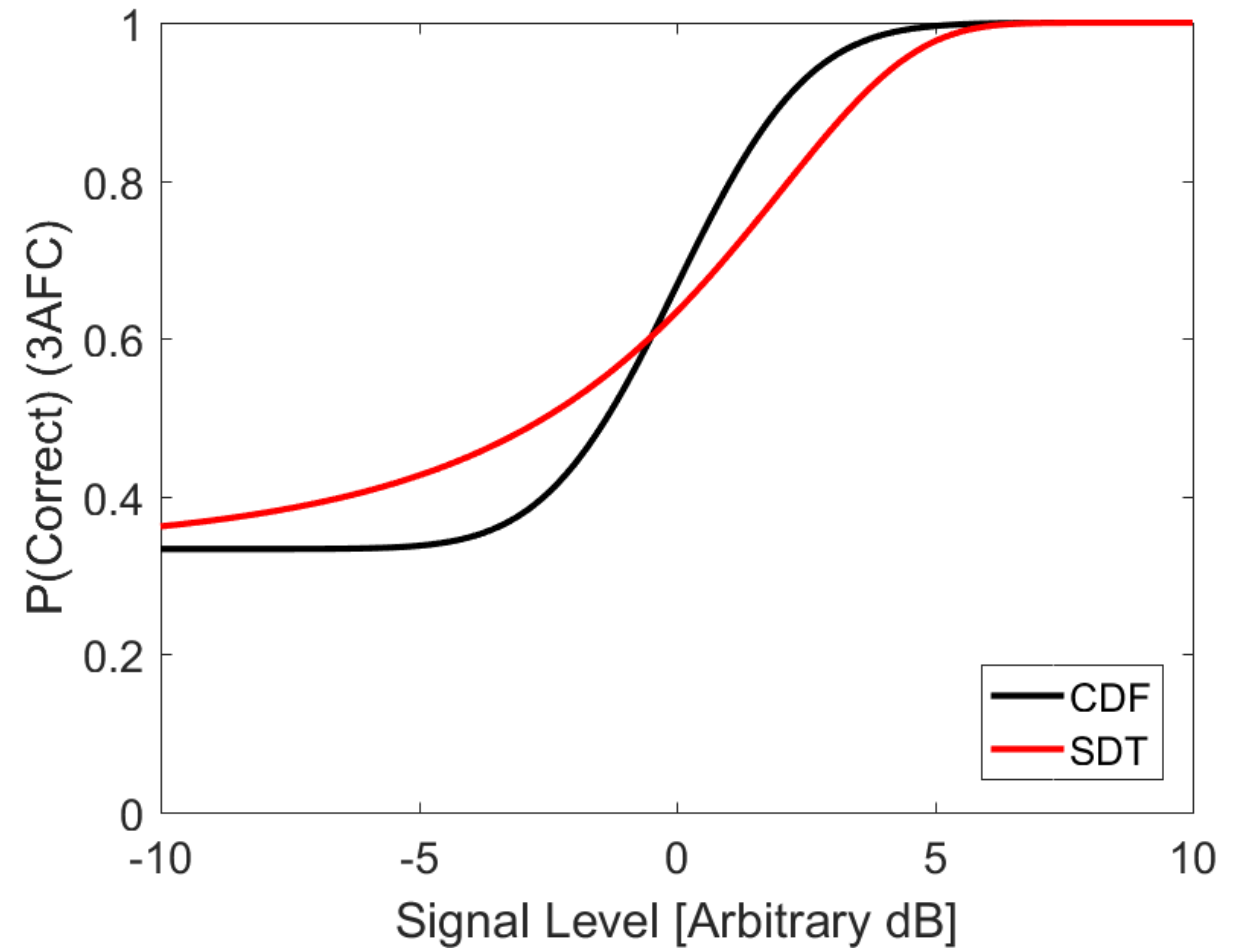
$$d' \propto S = 10^{\frac{dB}{10}}$$

$$P = \int_{-\infty}^{\infty} N_{PDF}(x - d') \cdot N_{CDF}(x)^{n-1} dx$$

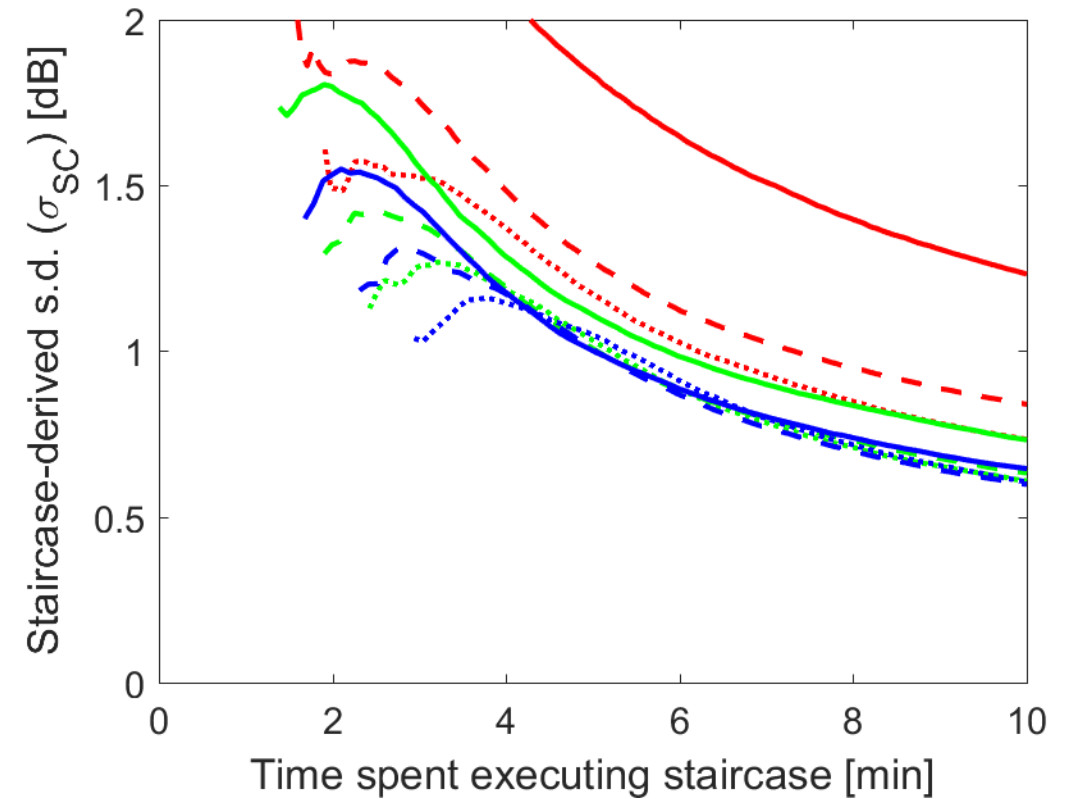
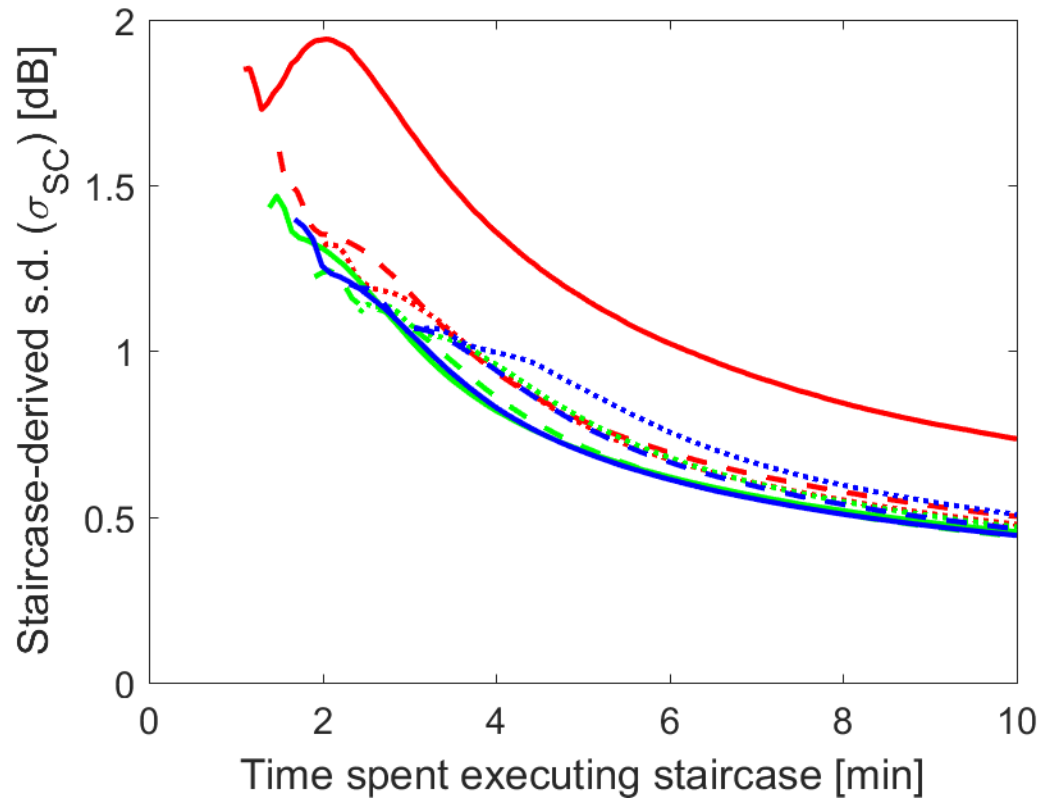


The SDT Psychometric Function

- Compared to the CDF, the SDT-derived function is considerably shallower in the region targeted by the 3D1U staircase ($P = .794$).
- This will lead to more scatter in the result in terms of dB.
 - The staircase method is actually estimating a location in terms of *probability*, so the scatter of the result is amplified by the inverse of the slope of the psychometric function.



SDT n and x Results



SDT Simulations



- The same staircase simulations can be run with the SDT function to evaluate this new scatter.
- It does not significantly change the result that 3AFC/3D1U is the best paradigm, so only these results are shown.
- The estimate of σ_{SC} changes +40%!

