

Jitter, Shimmer, and Noise in Pathological Voice Quality Perception

Jody Kreiman and Bruce R. Gerratt
UCLA School of Medicine



Introduction

- Jitter, shimmer, and harmonics-to-noise ratios are the cornerstones of acoustic voice measurement.
- The role these acoustic attributes play in the perception of voice quality remains unknown.
- Are listeners actually sensitive to differences in the amounts present within the range of clinical significance?

Experiment 1: Method

- 20 pathological voice samples (/a/, 1 s long)
- 10 male speakers, 10 female speakers

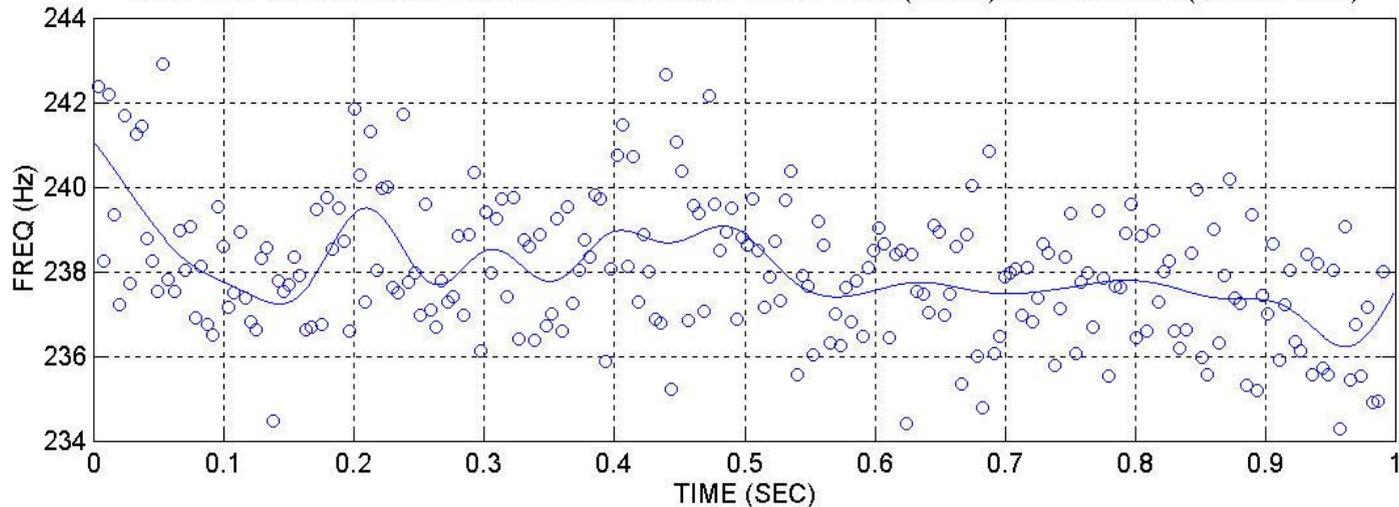
Analysis Methods

- Vocal tract resonance characteristics estimated via LPC.
- Cepstral comb liftering used to estimate the shape of the source noise spectrum.
- Characteristics of harmonic source estimated via inverse filtering and LF modeling.

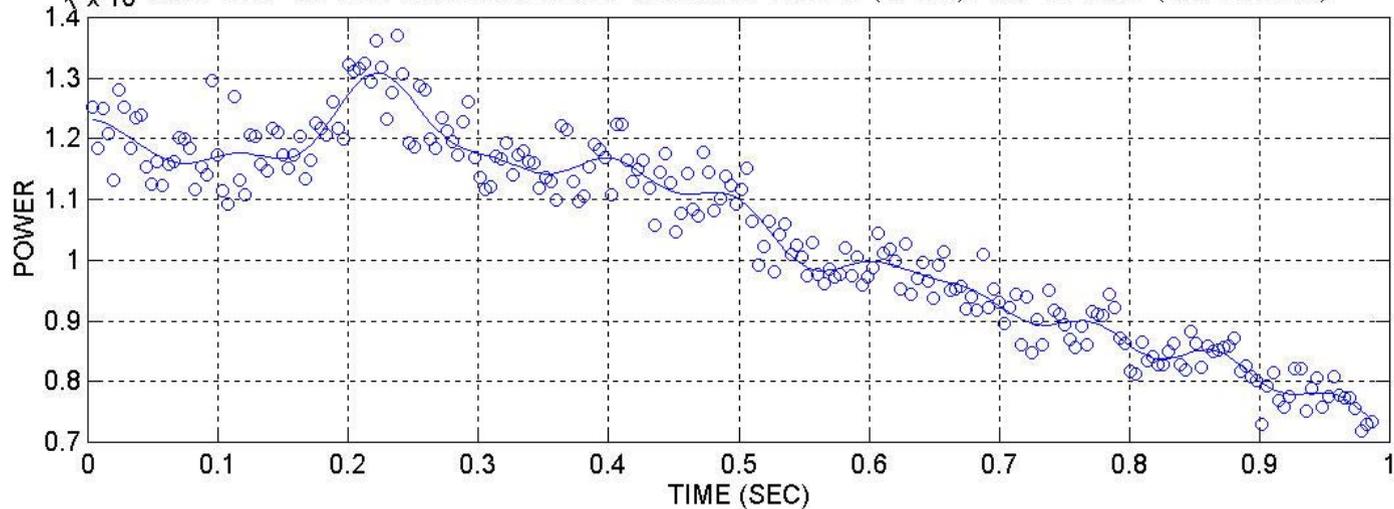
More Analysis Methods

- F0 tracked, and tremor rates estimated visually.
- Estimated tremor rate used as cutpoint between slow modulations (tremor) and fast modulations (jitter/shimmer).

9992. TRK=-31 CLK=2002112591257.74 ORIGINAL PITCH TRAJ (RED 0) AND TREMOR (GREEN LINE)



$\times 10^8$ 9992. TRK=-31 CLK=2002112591910.74 ORIGINAL POWER (RED 0) AND TREMOR (GREEN LINE)



Synthesis Methods

- Tremor modeled by incorporating low frequency amplitude and frequency tracks.
- Jitter was modeled by altering the duration of each cycle by an amount sampled from a high-pass filtered random sequence with $\sigma =$ desired level of jitter.
- Shimmer was modeled by altering the power of each cycle in a similar fashion.

More Synthesis Methods

- LF pulses were upsampled to 40 kHz and synthesized one by one.
- Frequency and amplitude of each pulse adjusted to reflect tremor, jitter, shimmer.
- Pulses concatenated and downsampled to 10 kHz.
- Spectrally-shaped noise time series added to create complete source time series.
- Source filtered through vocal tract model.

Last Synthesis Methods

- All parameters were adjusted to provide the best possible perceptual match to the original stimuli.
- A pilot experiment confirmed that synthetic stimuli were not distinguishable from the original samples at better than chance rates.

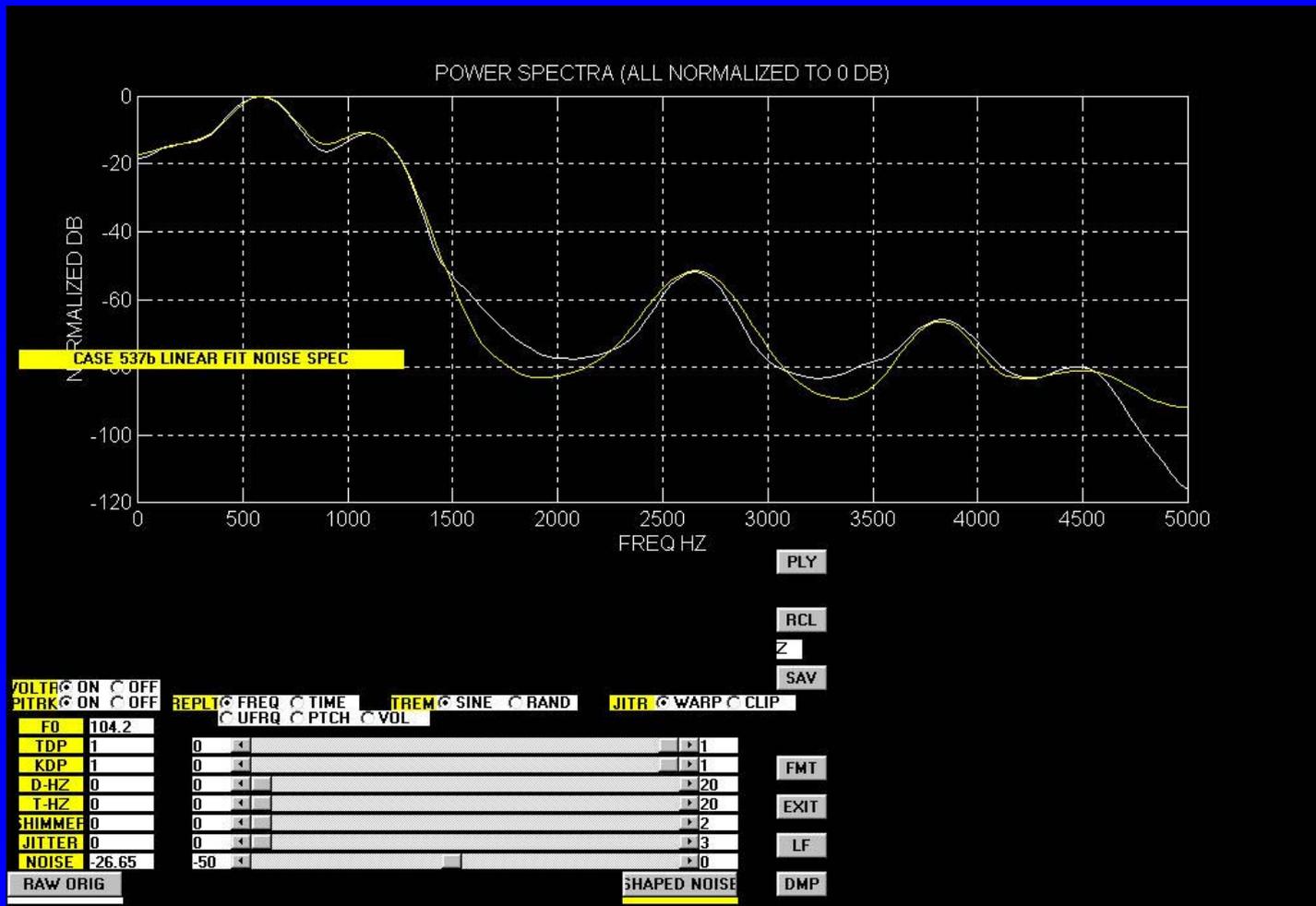
Perceptual Methods

- 70 naïve listeners, tested individually
- Listener Task: Adjust levels of jitter, shimmer, and/or noise to match the amount perceived in the natural voice sample.
- Parameters were adjusted singly and in combinations, in the context of spectral noise and without noise (7 total conditions).

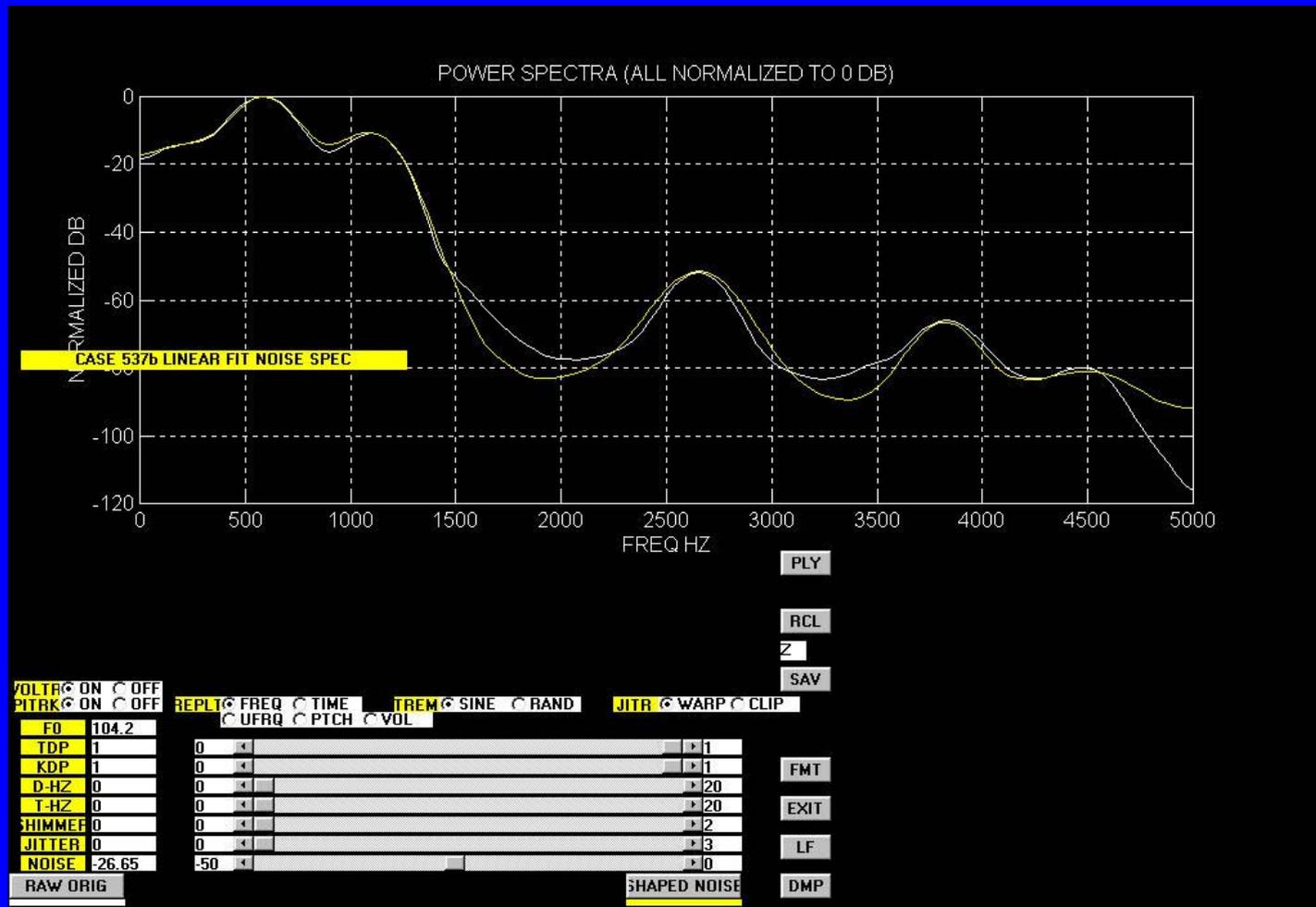
More Perceptual Methods

- Listeners heard a given voice once only, in a single condition.
- Ten listeners per voice, per condition.
- Two practice items were provided prior to testing.

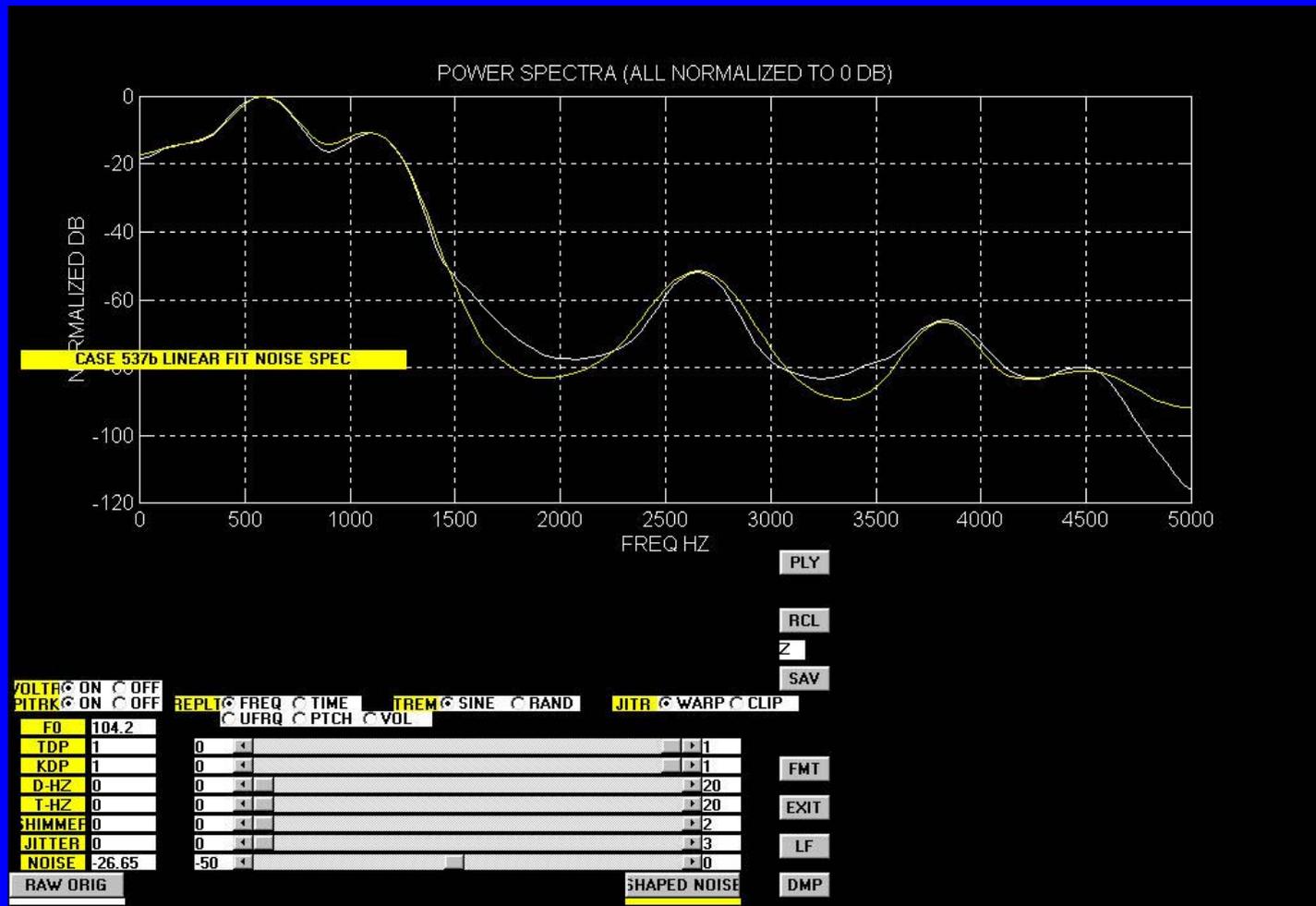
Jitter Adjustment Task



Shimmer Adjustment Task

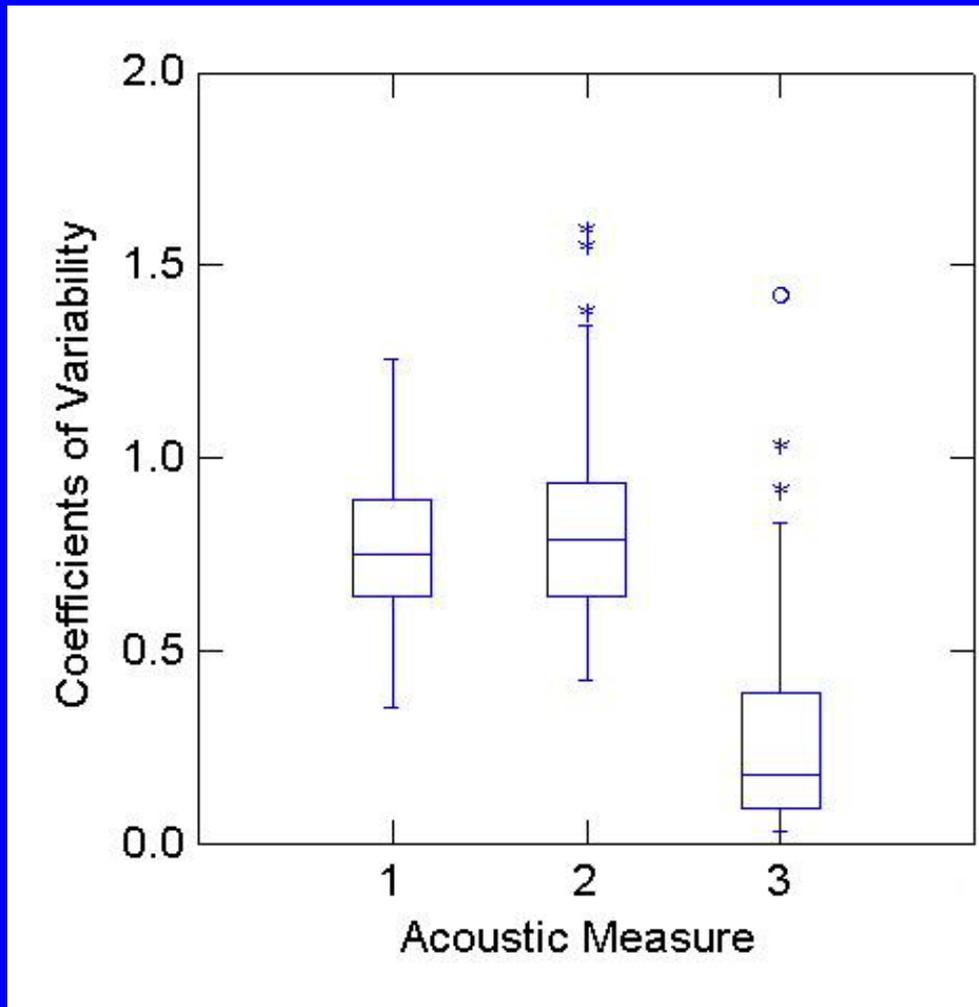


Noise Adjustment Task



Results

- Response variability measured with coefficients of variability.
- Listener responses were significantly more variable overall for jitter and shimmer than they were for noise.



1 = jitter; 2 = shimmer; 3 = noise

More Results

- Jitter and shimmer responses varied significantly with the listening task. Noise responses did not vary significantly across tasks.
- Variability in noise responses could be predicted in part by severity of deviation of the voice and by the shape of the harmonic source spectrum. Variability in jitter and shimmer responses were unpredictable.

Discussion

Two explanations for these results are possible.

- Listeners may be insensitive to differences in amounts of jitter and shimmer in a voice.
- Listeners may have difficulty determining which level is the correct response because they cannot separate jitter or shimmer perceptually from the composite noise component.

Experiment 2

- Eight voices selected
- For each voice, 5 series of 5 stimuli each were synthesized to embrace the range of listener responses from Experiment 1



Noise series



Jitter + noise series



Jitter only series



Shimmer + noise series



Shimmer only series



Perceptual Methods

- 18 listeners
- Pairs of stimuli
- Same/different task with confidence ratings
- ROC analysis

Results

- Listeners could not hear differences in jitter levels when stimuli included perceptually appropriate amounts of noise.
- Discrimination was better without noise, but sensitivity was still rather limited.

More Results

- Similar results occurred for shimmer. DLs in the context of spectral noise averaged about 1.3 dB, with a maximum of about 2 dB.
- Results for noise replicated our previous finding. Difference limens averaged about 10 dB, but vary substantially with the spectrum of the harmonic part of the source.

Discussion

- Listeners are remarkably insensitive to jitter and shimmer in natural voice contexts.
- Difference limens are large compared to values usually treated as clinically or experimentally meaningful.
- Listeners are more sensitive to overall amounts of spectral noise present.

More Discussion

- Given the long history of research on jitter and shimmer, we were surprised at how insensitive listeners are to these acoustic attributes.
- The question of the perceptual significance of any acoustic measure deserves primary attention during measurement development.
- Correlational studies can only describe associations, not show if or how a signal evokes a perceived quality.

Acknowledgments

- Synthesizer written by Brian Gabelman
- Additional programming support from Norma Antonanzas
- Jason Mallory created stimuli and tested listeners
- Research supported by NIH/NIDCD grant DC01797.