Reconciling quantification and meaning in voice quality (For Didier)

Jody Kreiman
Depts. of Head and Neck Surgery and Linguistics
UCLA

First things first

- Thank you so much for inviting me!
- Where I met Didier
- We continue to have overlapping interests, most recently in the biological aspects of voice.
- This talk: An exploration of some interdisciplinary ideas in the spirit of Didier’s career

What motivated this study?

- This project began with a simple question: *Why do we insist on describing voices as breathy and rough?*
  - Ample evidence has long indicated that:
    - Listeners do not agree in their ratings of breathiness and roughness.
    - Listeners do not even agree if a voice is breathy or rough.
  - Nevertheless, these terms (and others like them) are in constant use in both scientific and informal contexts.
  - If these terms are in fact meaningless, why have they persisted for millenia?
- Reconsideration appears to be in order.

Motivation and approach

- My approach today:
  - First examine attempts to connect meaning to sound, versus connecting sound to meaning.
  - We have not solved the problem of connecting voice signals to the meaning they convey to listeners, despite long-standing efforts.
  - Identify the limits of both descriptive and quantitative studies.
- Examine some new data that may show a way out of this age-old issue.

The dual nature of voice

- Two basic approaches to the study of voice quality: **Descriptive and quantitative**.
- From **a descriptive (and often humanistic) perspective**, voice quality resides in the listener, and not in the speaker.
- Listeners' affect and memory, the conversational setting, cultural structures, context, and other factors affect a voice's meaning.
- *In this tradition, what is heard and how it is understood do not depend solely on the physical signals, so meaning can never be fully explained or predicted by acoustic measures.*
- Consequent reliance on descriptive language, rather than measurement.

Examples: Author Raymond Chandler (1888-1958)

- Author Raymond Chandler (1888-1959) wrote remarkably about voices in works like *The Big Sleep* and *The Long Goodbye*.

- The meaning the protagonist derives from each voice is clear in Chandler's writing, but it is very difficult for readers to imagine what the voice actually sounds like.

Examples

“Please don’t get up,” she said in a voice like the stuff they use to line summer clouds with.

—*The Long Goodbye*

The voice I heard was an abrupt voice, but thick and clogged, as if it was being strained through a curtain or somebody’s long white beard.

—*The Little Sister*

He sounded like a man who had slept well and didn’t owe too much money.

—*The Big Sleep*

Quantifying signal meaning?

- Descriptive approaches can provide profound insight into what listeners hear when they listen to a voice.
- It is not possible to account for the detailed meanings a voice conveys via measurements of the signal, because these meanings inhere in part in the listener, and not solely in the speaker.
- Acoustic measures cannot provide a complete link between speakers and listeners—there is a gap in the “speech chain”.

Quantitative (scientific) approaches to voice quality

- Quantitative approaches to measuring voice quality assume that quality derives from speech production.
  - Inherent in the acoustic signal, so it is reasonable to expect listeners to agree in their ratings on scales.
- Quantitative methods have increased our understanding of the relationship between physical structures and processes and the sounds they produce.
  - Causally link sounds to the bodies that produced them.
- Methods generalize across studies and voices

Why quantification isn't the whole answer

- However, acoustic measures have not consistently explained even the simplest dimensions of vocal meaning, much less "strained through a long white beard."
  - Age
  - Gender
  - Emotion
  - Personal identity
- General failure to find consistent correlates for commonly-used terms
  - Many kinds of creaky voice, breathy voice
  - Default use of cepstral peak prominence as a measure of "overall quality"
  - Repeated tweaks to multivariate algorithms like AVQI

The Timbral Abyss

- It appears that there is no effective way to use acoustics to assess signal meaning, and there is no way to explain what listeners hear in acoustic terms.
- In other words, you can't get to meaning through acoustics, and you can't get to acoustics via meaning.
- Hence, we confront

THE TIMBRAL ABYSS,

a "conceptual and methodological barrier that prevents the reconciliation and integration of perspectives" (Van Elferen, 2017; Wallmark, 2022).

Van Elferen, I., Contemporary Music Review 36, 614 (2017); Wallmark, Z., Nothing but Noise (OUP, 2022)

The dilemma we face

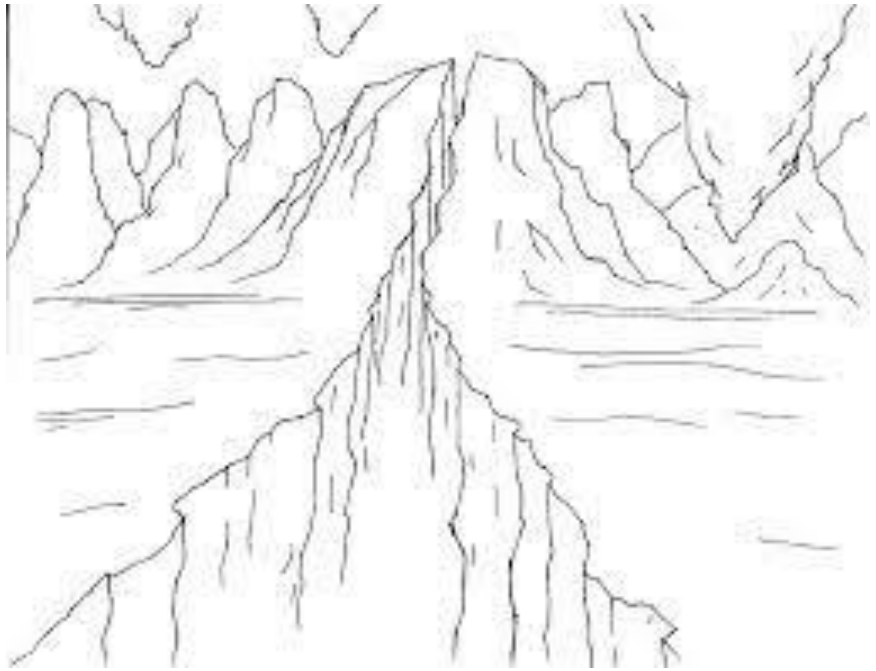
- Voice is produced as an acoustic signal, but…
  - what we hear is not a function solely of that signal.
  - what we describe is not explicable via instrumental measures.
- Thus, no one kind of analysis appears on its own to assess quality, and there is no obvious way to reconcile the different approaches, so that …
- ***the timbral abyss appears to be structural, and thus uncrossable*.**

BUT…

- Here is where reconsideration begins.
- Maybe there are points of connection between these modes of analysis, at least in some measure.
- Let’s back up for a moment and consider each research domain in a bit more detail.

State of the art on the two sides of the abyss

Describing



Quantifying

Part 1: Developing a vocabulary for quality

- Typical studies: Ratings on scales, free description, factor analysis
- Typical result: Across papers and meta-analyses a small set of dimensions consistently emerges from a limitless vocabulary.
  - Brightness/brilliance/ sharpness/clarity
    - Associated with the distribution of spectral energy in the voice
  - Breathiness and/or roughness
    - Associated with noise or spectral irregularity
  - Fullness/richness
    - Associated with the location of the spectral centroid

Examples

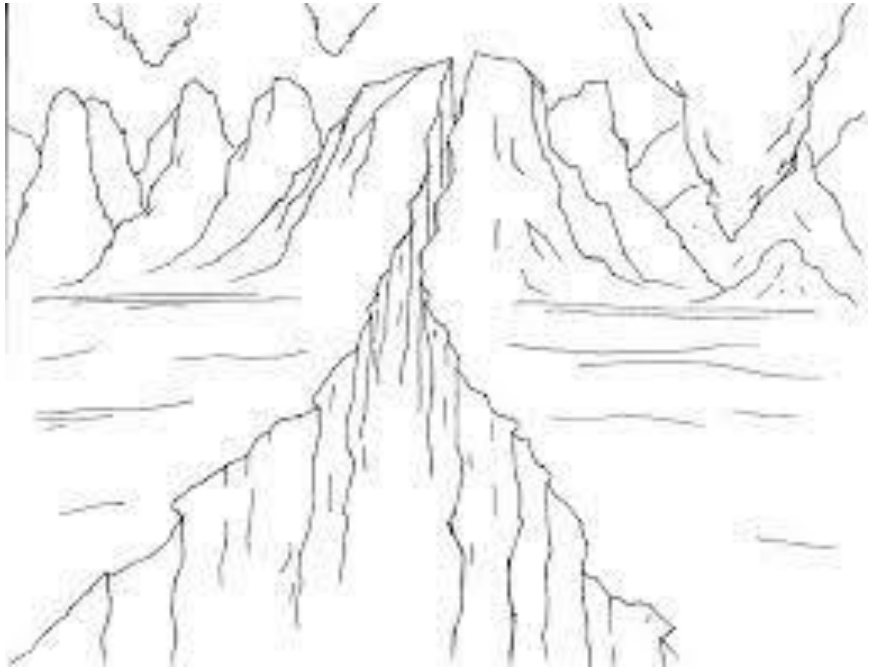
- Lichte, 1941
  - Paired comparisons, synthetic tones
  - Found evidence for at least 3 attributes: brightness, roughness, fullness
- Voiers, 1964
  - 16 male voices rated on 49 bipolar scales by 32 listeners
  - Factor analysis
  - 4 factors emerged: clear/hazy, roughness, magnitude, animation
- Pratt & Doak, 1976
  - 42 raters assessed usefulness of 19 common rating scales
  - 3 terms won the contest: brilliant/dull, warm/cold, rich/pure
- Similar sets of scales emerged across cultures and languages and from studies of instrumental timbre and animal vocalization.

Lichte, W.H., J. Exp. Psychol. 28, 455 (1941); Voiers, W.D., J. Acoust. Soc. Am 36, 1065 (1964); Pratt, R.L., & Doak, P.E., J. Sound Vibration 45, 317 (1976)

State of the art on the two sides of the abyss

Describing

1) Brightness/brilliance/ sharpness/clarity
2) Breathiness and/or roughness
3) Fullness/richness



Quantifying

Part 2: Relating acoustic measures to meaning

- The “rate and correlate” approach
  - Seeks acoustic cues to individual qualities or characteristics (breathiness, age, sex…)
  - Often multivariate
  - Typically use correlation or regression to associate terminology with measurements
  - Results are highly variable across studies

Relating acoustic measures to meaning

- Studies of the acoustic attributes that distinguish different voices (Lee & Kreiman)
  - Basic idea: variability = meaning
  - Acoustic analyses of many, many voices; lots of different kinds of samples.
  - Principal component analysis to determine dimensions of acoustic voice space
- Largest variance components weighed on *the balance of high-frequency harmonic and inharmonic energy* in the voice (related to strained/breathy continuum) and on *formant dispersion*.

Relating acoustic measures to meaning

- Variations in the balance of harmonic and inharmonic energy in the voice source
  - Often associated with a quality continuum from “strained” or “pressed” (or “bright”) to “breathy”
  - Signals arousal across many species
- Formant dispersion
  - Related to location of spectral centroid/balance of spectral energy
  - Signals both dominance and reproductive fitness across many species

Relating acoustic measures to meaning

- Same result for virtually every speaker regardless of age, sex, and language spoken.
  - Most likely explanation: derived through evolution
  - Additional components reflect language characteristics (tone, phonation contrasts) and idiosyncracies of individual talkers.
- These factors also characterize vocalization across many species and provide survival benefits.
  - Reproductive fitness
  - Hostile/benign intent
  - Physical size
  - Arousal
- ***Part of the biological purpose of phonation, and hence its meaning***

Physiological control of voice quality

- Zhaoyan Zhang's work suggests that speakers control voice quality primarily by manipulating vocal fold medial thickness, complemented by vocal tract adjustments.
- Further evidence of evolutionary origins.
  - Thicker folds → better airway protection (basic function of the larynx)
  - Many acoustic measures are also related to vocal fold thickness.
    - H1-H2, H1-H2kHz, cepstral peak prominence

Zhang, Z., J. Voice, in press (2023).

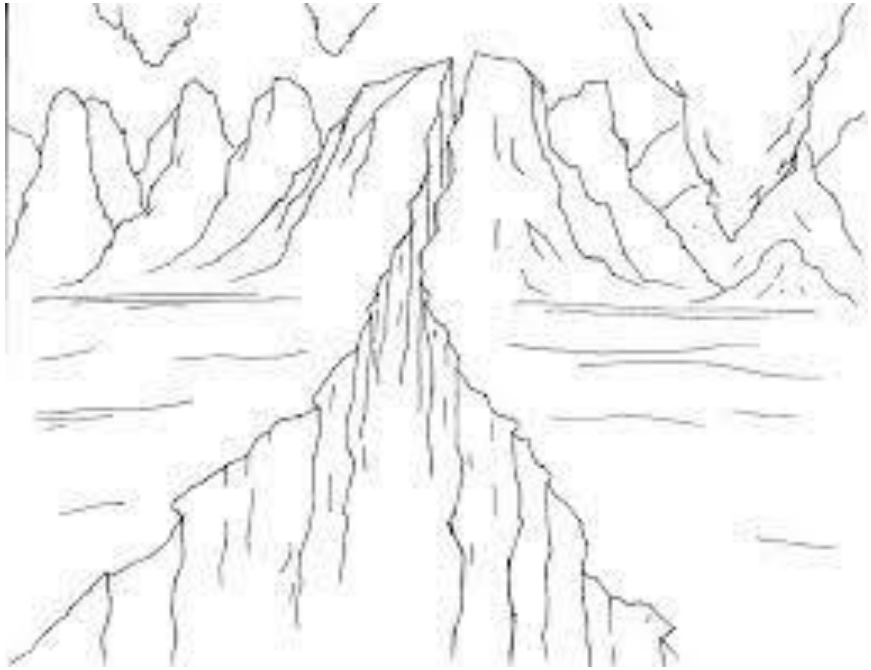
Crossing the abyss

- Empirical evidence points to universal factors that establish a voice space that applies to all voices.
- The dimensions of this space strongly resemble those that commonly emerge from descriptive studies of voice.

State of the art on the two sides of the abyss

Describing

1) Brightness :: distr of spectral energy
2) Breathiness :: noise
3) Fullness/richness :: spectral centroid



Quantifying

1) Harmonic/inharmonic energy :: breathy to pressed
2) Formant dispersion /spectral centroid :: richness/fullness

Why these scales, and not others?

- We use these terms, and not others, because they have biological relevance, apply to every voice, and thus carry a consistent meaning regardless of who is talking or who is listening.
- Acoustics and semantics converge at this biological point.
- Wallmark & Kendall (2018) suggested a similar association.
  - Recurring descriptors recur because they link voices to bodies.
- The present account takes this further: WHY this SPECIFIC SET of descriptors plays this role.

The timbral abyss is not a bottomless pit

- This correspondence between the most common terms for voice and parameters that define the human acoustic voice space suggests that the meaning voices carry rests on a bedrock of biology.

- “Breathy” and “rough” remain useful
  - because they directly link bodies and signals to perceived voices, and to meaning, seemingly without the need to consider the specific perceptual, cognitive, or emotional context surrounding the act of hearing, and
  - because they reliably carry (seemingly) universal meanings.

Limitations

- There is much more to voice acoustics than just these few shared dimensions.
- The vocabulary available to describe or discuss voices is essentially limitless.
- ***It is thus likely that a model that completely maps from one domain to another is both theoretically and empirically impossible.***

Limitations to the limitations

- Nevertheless, this foundation provides a way of explaining a few critical facets of the meaning of voice in terms of specific details of production and acoustics, and vice versa, ***thus spanning, at least in part, the timbral abyss.***
- ***Not the case that we can't connect form and function in voice—it's just that the connection is small compared to the amount of work voice does in conveying information.***
- Exploiting this bridge in our ongoing work could help integrate studies of physical signals and their meaning, leading eventually to a truly interdisciplinary approach to voice.

Parting words

Then she laughed. It was almost a racking laugh. It shook her as the wind shakes a tree. I thought there was puzzlement in it, not exactly surprise, but as if a new idea had been added to something already known and it didn't fit. Then I thought that was too much to get out of a laugh.

— Raymond Chandler, *The Big Sleep*

Thank you!