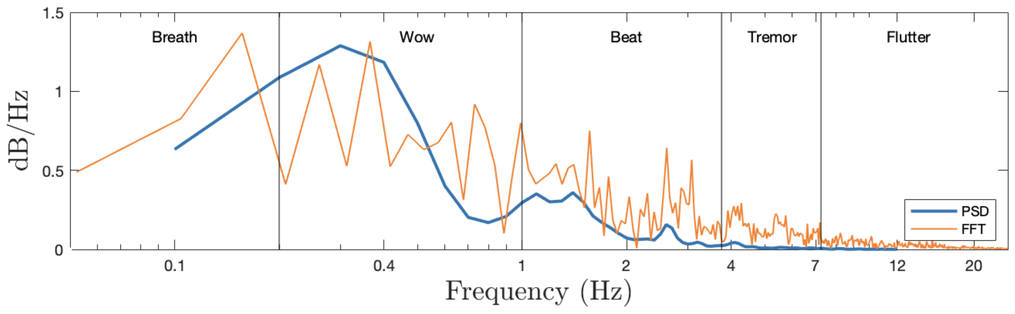
Sources and interactions of long-term phonatory instabilities

François-Xavier Brajot.

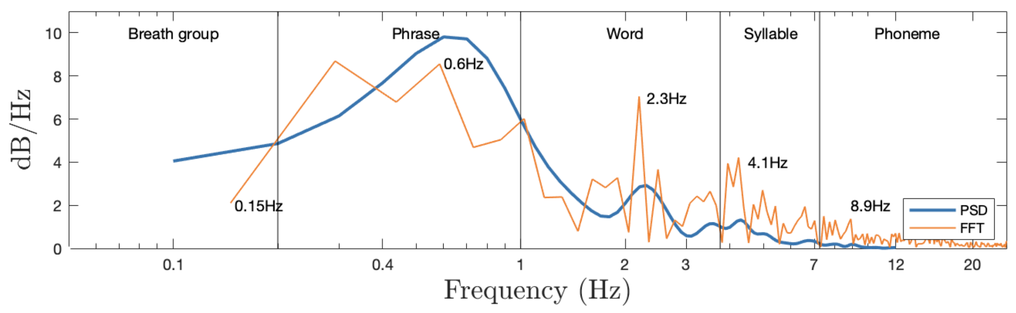
Why all the noise?

- Slow fluctuations in voice f_o and intensity have been coarsely distinguished with respect to modulation frequency.
  Three have presumed neurologic origins associated with auditory feedback (wow), somatosensory feedback (tremor), and motor unit discharge (flutter).
- These modulations account for much of the variability in f_o (i.e. ⅔ of variance accounted for by lower 4 Hz, ⅓ just by wow alone).
  Wow and tremor can be characterized by non-linear dynamics of negative feedback loops and are amenable to sensory perturbations.
- There is an important overlap between long-term phonatory instabilities and major speech cycles (breath group, phrase, word, syllable, phoneme).
  Speech units may therefore be biologically derived, constrained, or influenced by the same mechanisms that give rise to long-term phonatory instabilities.

Spectrum of vocal f_o during sustained phonation.



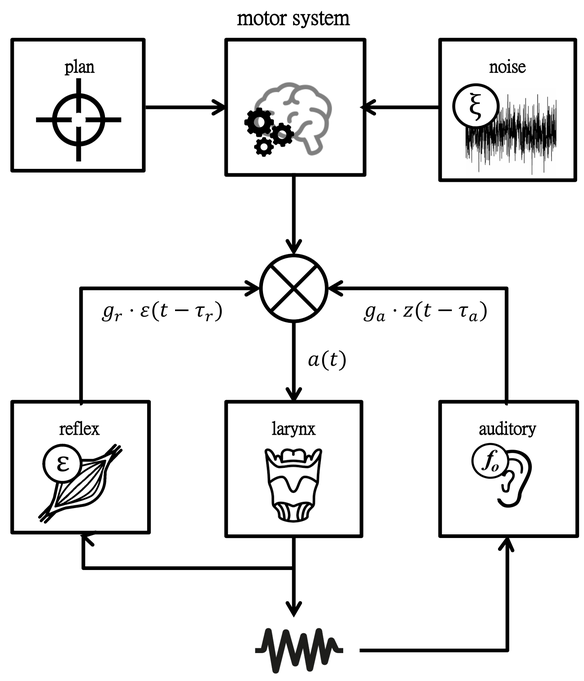
Spectrum of vocal f_o during sentence production.



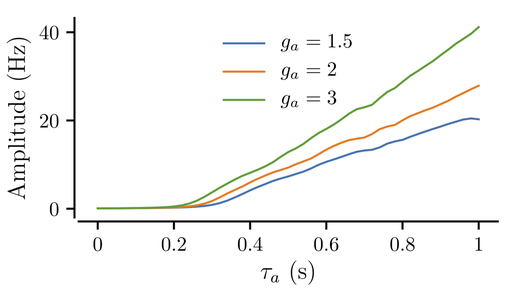
Model

In the reflex resonance model proposed by Titze et al. 2002 JASA 111: 2272, laryngeal muscle reflexes are mediated by feedback gain g_r and delay τ_r.
Increasing g_r results in a 4-7Hz vibrato (tremor).

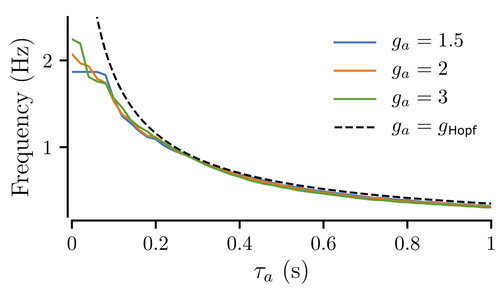
We incorporated an auditory feedback loop.
In the linear case, the system is overdamped; oscillations are driven by noise ξ.
With non-linearity in auditory feedback, self-sustained 0.2-2Hz wow oscillations appear for critical values of auditory gain g_a and delay τ_a (i.e. Hopf bifurcation).



Feedback gain → modulation depth

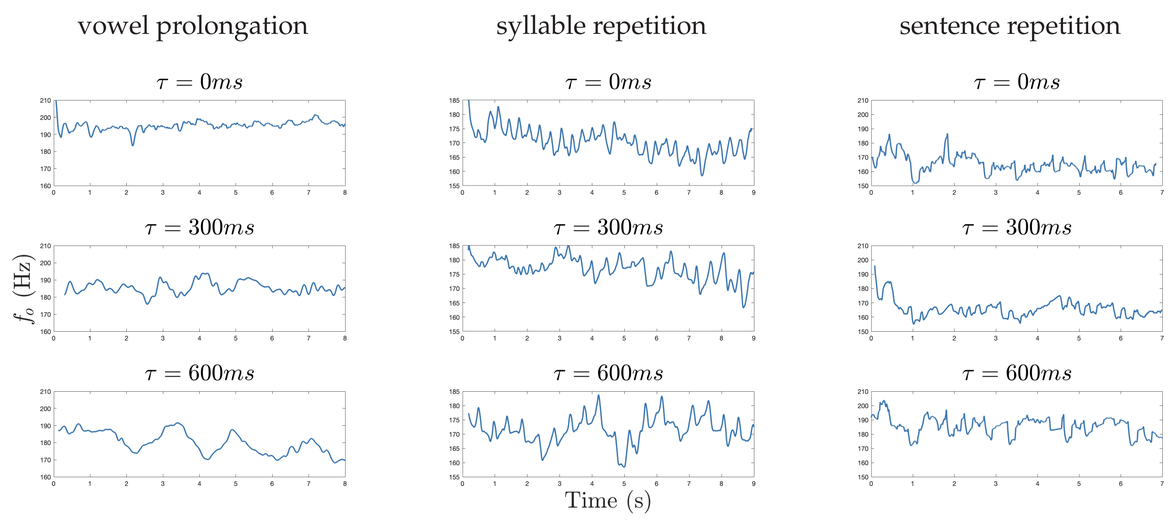


Feedback delay → modulation frequency



Vocal wow

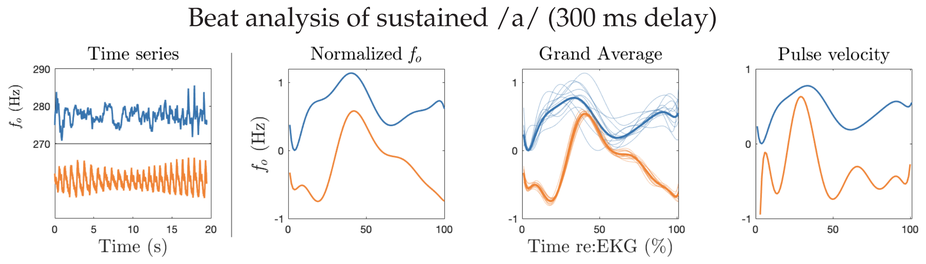
Delaying auditory feedback induces 0.2-2Hz modulations in vocal f_o.
This is consistent with the view that speech and voice use negative feedback.
This effect is visible in syllable repetition, but lost or hidden in connected speech.



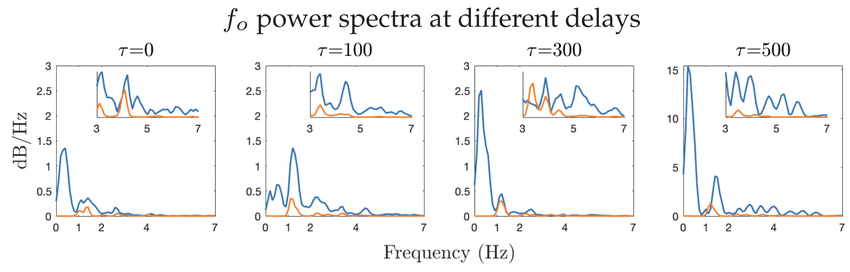
Vocal beat

A potential confound is the effect of blood flow on stiffness of laryngeal muscles.
Pulse systolic peak and "harmonics" extend 1-5Hz, possibly as high as 8Hz.

Beat analysis of sustained /a/ (300 ms delay)



f_o power spectra at different delays

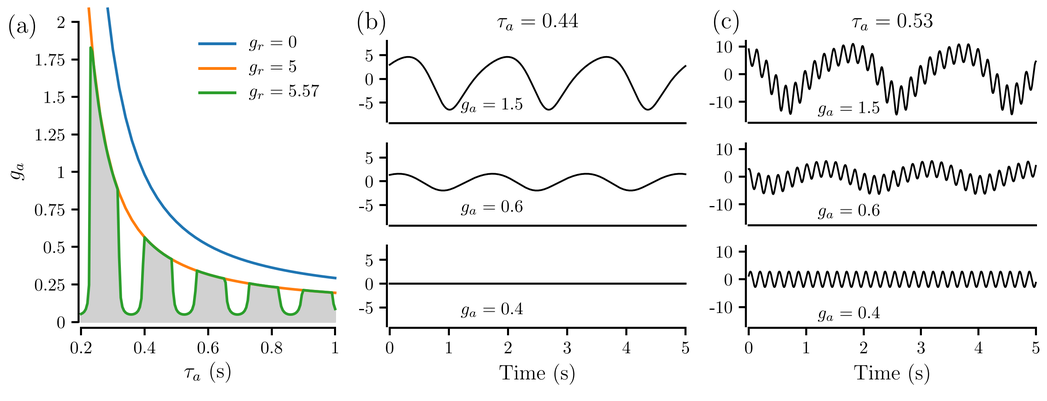


Interactions

Increasing reflex gain g_r lowers the threshold at which auditory feedback parameters reach Hopf bifurcation and stable limit cycle (blue → orange line).

As reflex gain approaches the bifurcation threshold for tremor, the relationship becomes non-monotonic (green line). Within the grey areas, tremor is suppressed.

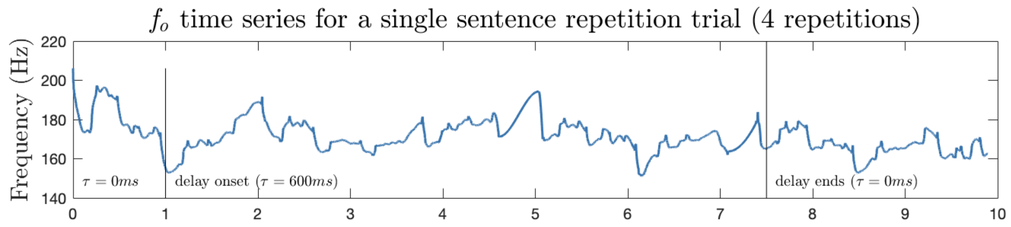
Hopf bifurcation lines and example f_o for given g_a and τ_a at $g_r = 5.57$



Demodulation

The modulations are time-varying, non-stationary and interact.
Empirical mode decomposition provides a way to extract "intrinsic modes".
There appear to be correlations between wow and phrase variations, beat and word-level variations - but this has yet to be properly tested experimentally.

f_o time series for a single sentence repetition trial (4 repetitions)



Largest intrinsic mode functions from EMD