Variational blind audio deconvolution with a Gaussian process kernel bank

Marnix Van Soom

VUB AI Lab / UA InViLab

2026-06-10

BIASLab, TU/e Eindhoven

1 An inverse problem

$$x(t) = e(t) * h(t)$$

$$\begin{aligned} p(x, e, h) &= p(x \mid e, h)\, p(e)\, p(h) \\ &= p(e, h \mid x)\, p(x) \end{aligned}$$

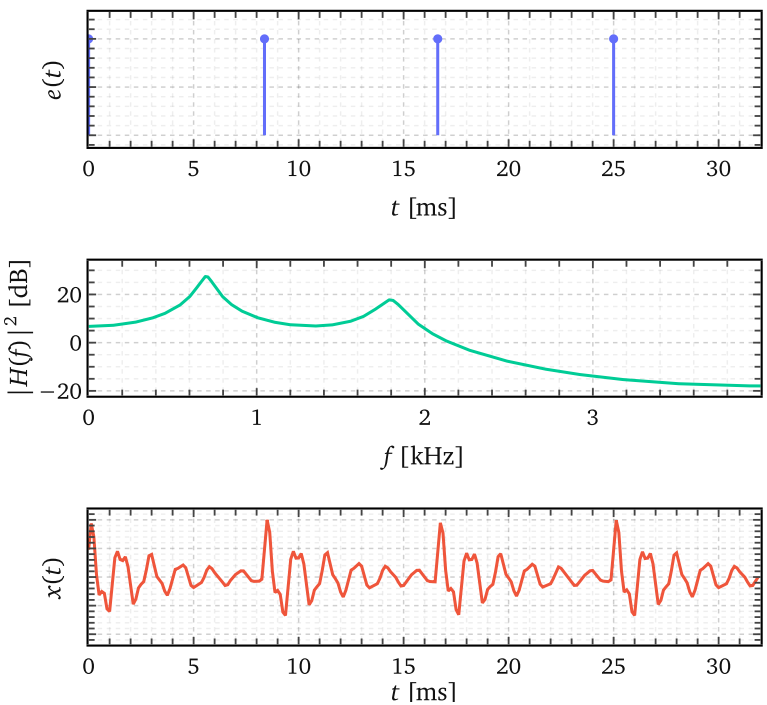
Blind audio deconvolution

$$x(t) = e(t) * h(t)$$
$$X(f) = E(f)\,H(f)$$

- Two directions:

$$(e, h) \longrightarrow (x) \qquad \text{forward}$$
$$(x) \longrightarrow (e, h) \qquad \text{inverse}$$

- Inverse is underdetermined
  - needs regularization
- Bayesian regularization
  - $p(x, e, h) = p(x \mid e, h)\,p(e)\,p(h)$



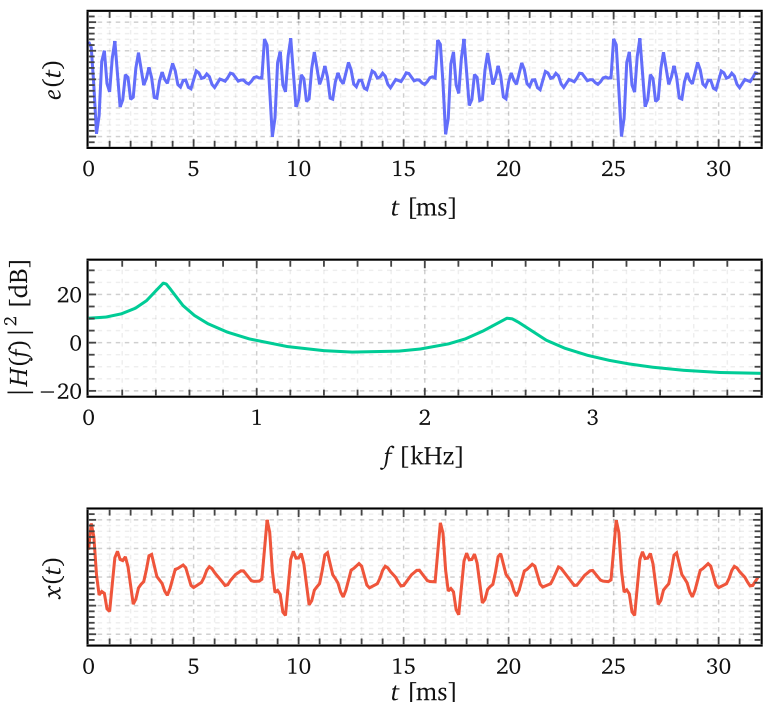
Blind audio deconvolution

$$x(t) = e(t) * h(t)$$
$$X(f) = E(f)\,H(f)$$

- Two directions:

$$(e, h) \longrightarrow (x) \qquad \text{forward}$$
$$(x) \longrightarrow (e, h) \qquad \text{inverse}$$

- Inverse is underdetermined
  - needs regularization
- Bayesian regularization
  - $p(x, e, h) = p(x \mid e, h)\,p(e)\,p(h)$

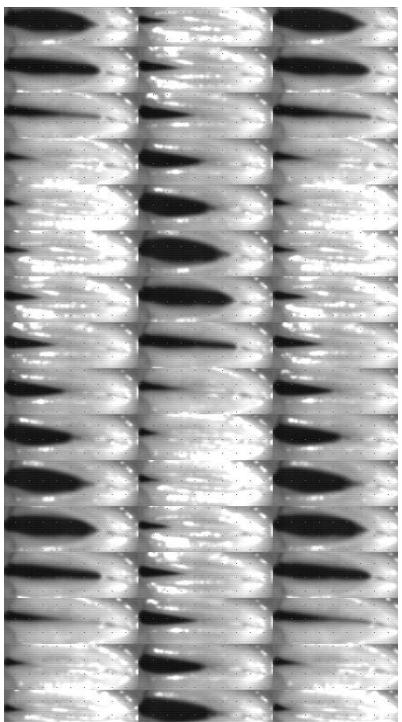


Glottal inverse filtering

$$x(t) = e(t) * h(t)$$
$$X(f) = E(f)\,H(f)$$

- $x(t)$: speech audio
- $e(t)$: glottal excitation

- $H(f)$: vocal tract filter



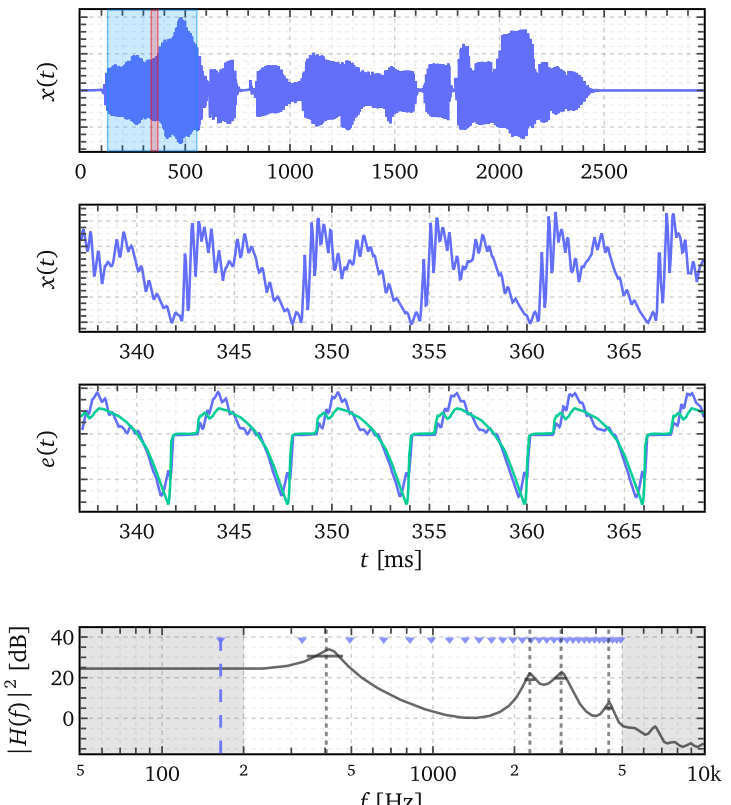
[Jopling 2009]

Glottal inverse filtering

$$x(t) = e(t) * h(t)$$

$$X(f) = E(f)\,H(f)$$

- $x(t)$: speech audio
- $e(t)$: glottal excitation
  - voiced/unvoiced
  - fundamental frequency F_0
  - jitter
- $H(f)$: vocal tract filter
  - formants F_1, F_2, F_3



2 Speech model

$$x(t) = e(t) * h(t)$$

$$\begin{aligned} p(x, e, h) &= p(x \mid e, h)\, p(e)\, p(h) \\ &= p(e, h \mid x)\, p(x) \end{aligned}$$

Linear prediction

- All-pole filter [Atal & Hanauer 1971]

$$H(z) = 1/A(z)$$

$$A(z) = 1 - a_1 z^{-1} - \cdots - a_P z^{-P}$$

- AR(P) model

$$x_m = \sum_{p=1}^{P} a_p \, x_{m-p} + e_m$$

- Forward:
  - $\mathbf{x} = \mathbf{\Psi}(\mathbf{a})^{-1}\mathbf{e}$
- Inverse:
  - $\mathbf{x} = \mathbf{X}\mathbf{a} + \mathbf{e}$

$$\Leftrightarrow (\mathbf{X}^\top \mathbf{X})\,\hat{\mathbf{a}} = \mathbf{X}^\top \mathbf{x}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ -a_1 & 1 & & & \\ -a_2 & -a_1 & 1 & & \\ & -a_2 & -a_1 & 1 & \\ & & -a_2 & -a_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} x_0 & x_{-1} \\ x_1 & x_0 \\ x_2 & x_1 \\ x_3 & x_2 \\ x_4 & x_3 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{pmatrix}$$

Kernel linear prediction

- Excitation model:

$$\mathbf{e} = \mathbf{u}' + \mathbf{n}$$

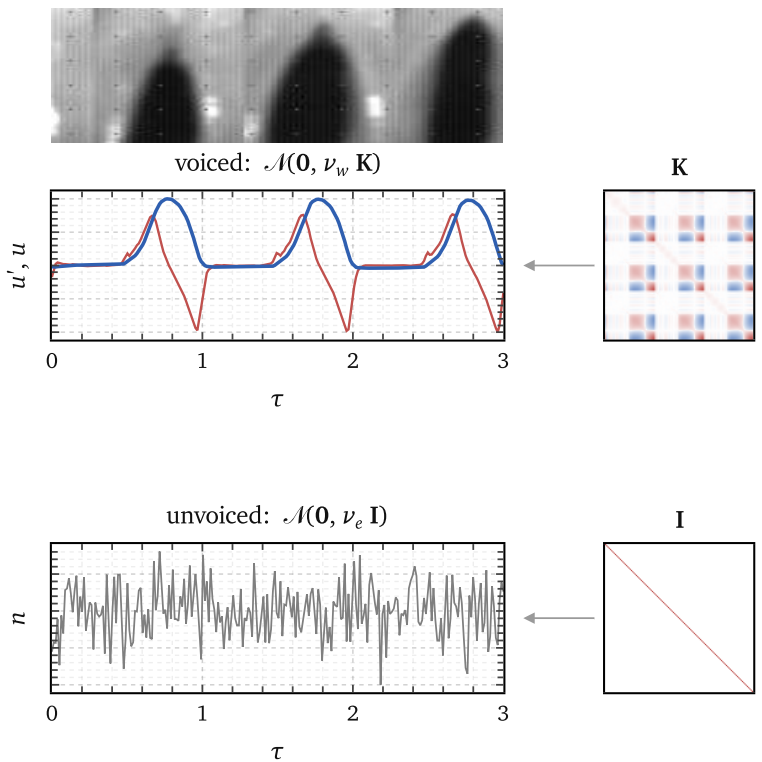
$$\mathbf{u}' \sim \mathcal{N}(\mathbf{0}, \nu_w \mathbf{K}) \qquad \text{voiced}$$

$$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \nu_e \mathbf{I}) \qquad \text{unvoiced}$$

- Kernelize!
  - model

$$u'(t) \sim \mathcal{GP}(0, k(t, t'))$$

  - design k for closure condition, periodicity, roughness, changepoints



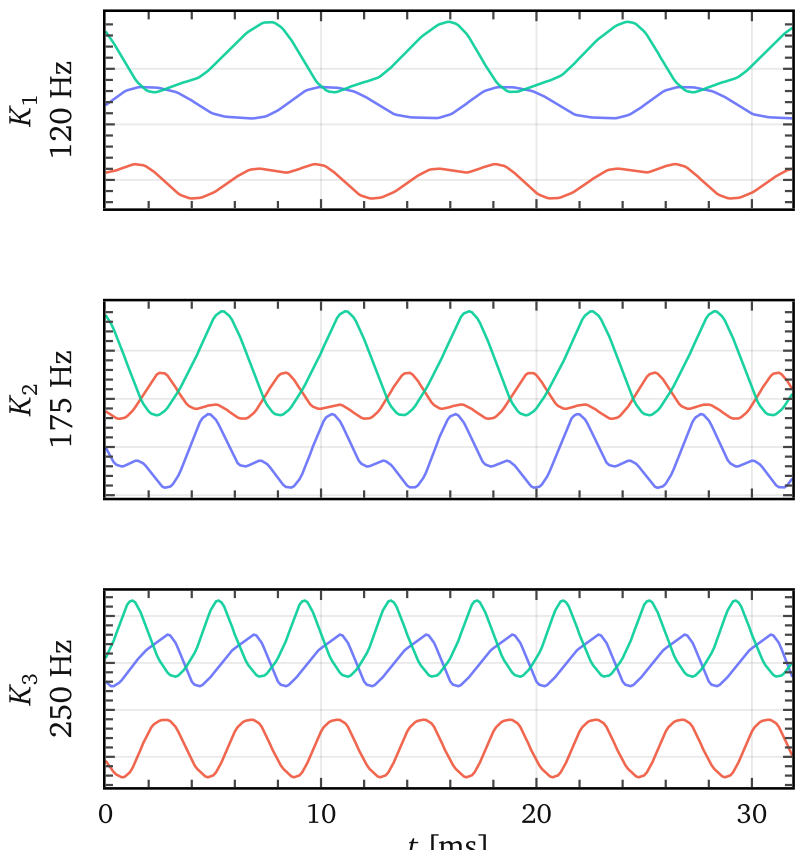
Kernel bank linear prediction

Speech model of [Yoshii & Goto 2013]:

- Kernel bank

$$\mathbf{K} = \sum_{i=1}^{I} \theta_i \mathbf{K}_i, \quad \theta_i \sim \text{Gamma}(\alpha/I, \alpha)$$

- Each $\mathbf{K}_i$ a periodic kernel
  - one pitch hypothesis
- I large: a few θ_i dominate
  - sparse selection



Variational inference

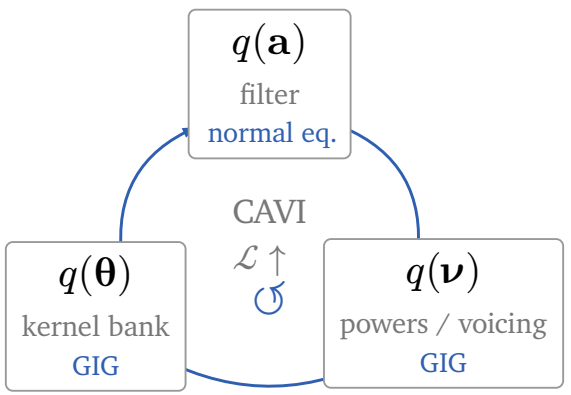
- Forward:

$$\begin{aligned}
\mathbf{a} &\sim \mathcal{N}(\mathbf{0}, \lambda \mathbf{I}) \\
\nu_w &\sim \text{Gamma}(a_w, b_w) \\
\nu_e &\sim \text{Gamma}(a_e, b_e) \\
\theta_i &\sim \text{Gamma}(\alpha/I, \alpha), \quad i = 1, ..., I \\
\mathbf{K} &= \nu_w \sum_i \theta_i \mathbf{K}_i + \nu_e \mathbf{I}
\end{aligned}$$

$$\mathbf{x} \mid \mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}(\mathbf{a})^{-1} \, \mathbf{K} \, \boldsymbol{\Psi}(\mathbf{a})^{-\top})$$

- Inverse:

$$\begin{aligned}
&\text{argmin}_q \, D_{\text{KL}}(q(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu}) \,\|\, p(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu} \mid \mathbf{x})) \\
&= \text{argmax}_q \, \mathcal{L}(q(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu}))
\end{aligned}$$



$$q(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu}) = q(\mathbf{a}) \, q(\boldsymbol{\theta}) \, q(\boldsymbol{\nu})$$

$$\mathcal{L} = \underbrace{\mathbb{E}_q[\log p(\mathbf{x} \mid \mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu})]}_{\text{fit}} - \underbrace{D_{\text{KL}}(q \,\|\, p(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu}))}_{\text{stay near prior}}$$

[Yoshii & Goto 2013]

Variational inference

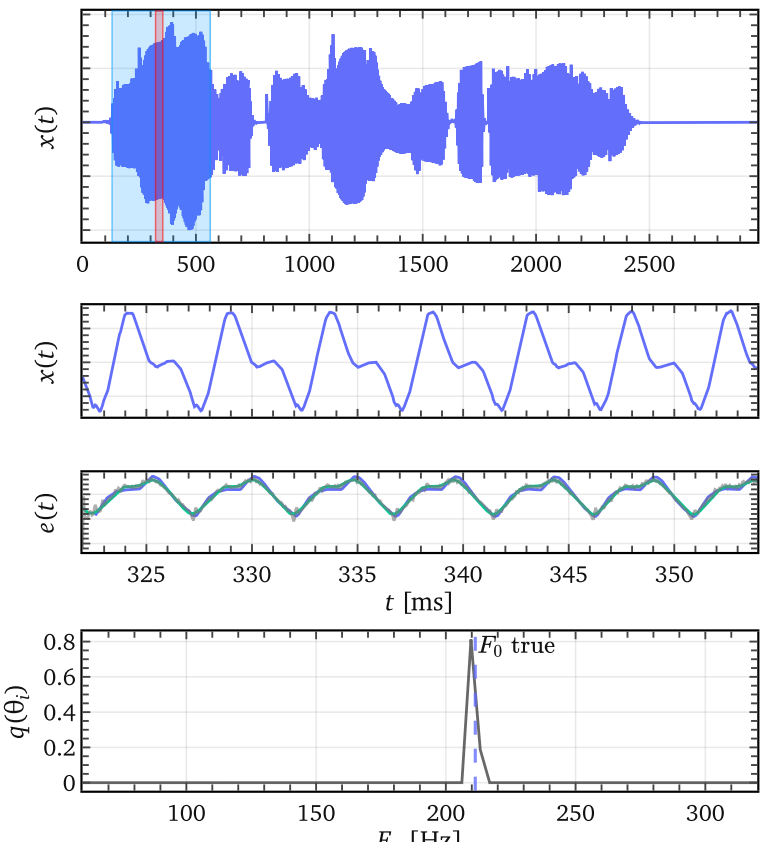
- Forward:

$$\begin{aligned}
\mathbf{a} &\sim \mathcal{N}(\mathbf{0}, \lambda \mathbf{I}) \\
\nu_w &\sim \text{Gamma}(a_w, b_w) \\
\nu_e &\sim \text{Gamma}(a_e, b_e) \\
\theta_i &\sim \text{Gamma}(\alpha/I, \alpha), \quad i = 1, ..., I \\
\mathbf{K} &= \nu_w \sum_i \theta_i \mathbf{K}_i + \nu_e \mathbf{I}
\end{aligned}$$

$$\mathbf{x} \mid \mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Psi}(\mathbf{a})^{-1} \mathbf{K} \boldsymbol{\Psi}(\mathbf{a})^{-\top}\right)$$

- Inverse:

$$\begin{aligned}
&\text{argmin}_q \, D_{\text{KL}}(q(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu}) \,\|\, p(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu} \mid \mathbf{x})) \\
&= \text{argmax}_q \, \mathcal{L}(q(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu}))
\end{aligned}$$



Variational inference

- Forward:

$$\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \lambda \mathbf{I})$$
$$\nu_w \sim \text{Gamma}(a_w, b_w)$$
$$\nu_e \sim \text{Gamma}(a_e, b_e)$$
$$\theta_i \sim \text{Gamma}(\alpha/I, \alpha), \quad i = 1, ..., I$$
$$\mathbf{K} = \nu_w \sum_i \theta_i \mathbf{K}_i + \nu_e \mathbf{I}$$
$$\mathbf{x} \mid \mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}(\mathbf{a})^{-1} \mathbf{K} \boldsymbol{\Psi}(\mathbf{a})^{-\top})$$

- Inverse:

$$\text{argmin}_q D_{\text{KL}}(q(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu}) \| p(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu} \mid \mathbf{x}))$$
$$= \text{argmax}_q \mathcal{L}(q(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\nu}))$$

BNGIF
Bayesian nonparametric glottal inverse filtering
=

kernel bank linear prediction $p(x, e, h)$

model and inference from [Yoshii & Goto 2013]

+

informative AR prior $p(h)$

$p(\mathbf{a} \mid \text{formants})$ from AR theory

+

data-driven kernel bank $p(e)$

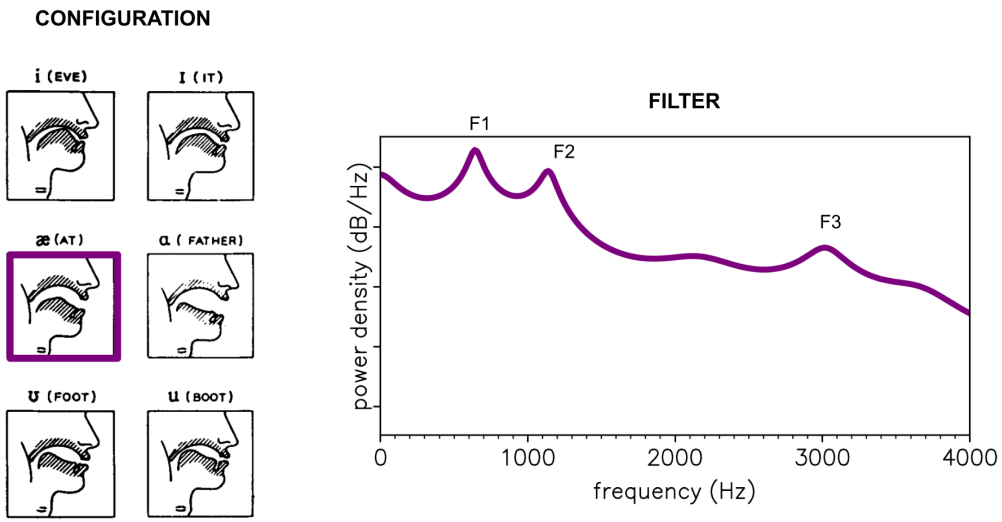
$\{\mathbf{K}_i\}$ learned from large synthetic corpus

3 Filter prior

$$x(t) = e(t) * h(t)$$

$$\begin{aligned} p(x, e, h) &= p(x \mid e, h)\, p(e)\, \boldsymbol{p(h)} \\ &= p(e, h \mid x)\, p(x) \end{aligned}$$

Vocal tract filters



[Rabiner et al.1993]

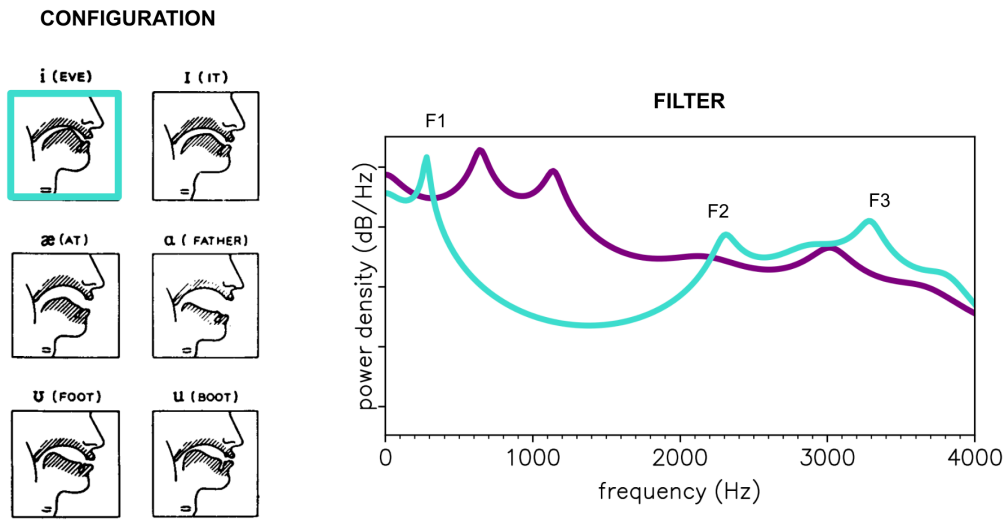
$$H(z) = 1/A(z)$$

$$A(z) = 1 - \sum_{p=1}^{P} a_p z^{-p}$$

$$= 1 - 1.56z^{-1} + 2.29z^{-2} - 2.40z^{-3} + 2.12z^{-4} - 1.31z^{-5} + 0.81z^{-6}$$

$$= (1 - 1.70z^{-1} + 0.96z^{-2})(1 - 0.46z^{-1} + 0.93z^{-2})(1 + 0.59z^{-1} + 0.91z^{-2})$$

Vocal tract filters



[Rabiner et al.1993]

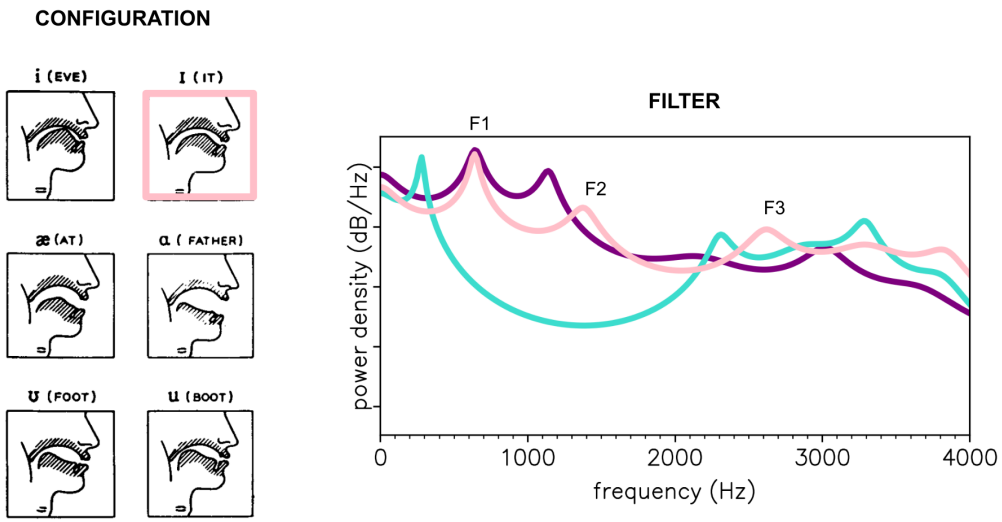
$$H(z) = 1/A(z)$$

$$A(z) = 1 - \sum_{p=1}^{P} a_p z^{-p}$$

$$= 1 - 0.12z^{-1} - 0.04z^{-2} - 1.28z^{-3} - 0.01z^{-4} - 0.03z^{-5} + 0.82z^{-6}$$

$$= (1 - 1.92z^{-1} + 0.96z^{-2})(1 + 0.44z^{-1} + 0.94z^{-2})(1 + 1.36z^{-1} + 0.91z^{-2})$$

Vocal tract filters



[Rabiner et al.1993]

$$H(z) = 1/A(z)$$

$$A(z) = 1 - \sum_{p=1}^{P} a_p z^{-p}$$

$$= 1 - 1.08z^{-1} + 1.33z^{-2} - 1.92z^{-3} + 1.24z^{-4} - 0.88z^{-5} + 0.81z^{-6}$$

$$= (1 - 1.86z^{-1} + 0.96z^{-2})(1 - 0.02z^{-1} + 0.93z^{-2})(1 + 0.80z^{-1} + 0.91z^{-2})$$

Spectral features

- A spectral feature is a polynomial
  - $Q(z) = 1 - 1.69\, z^{-1} + 0.95\, z^{-2}$
  - $Q(z) \mid A(z)$
- K spectral features
  - $M(z) = \mathrm{lcm}\{Q_1, ..., Q_K\}$
  - $M(z) \mid A(z)$
- Observation:
  - $M \mid A \iff \mathbf{Fa} + \mathbf{c} = \mathbf{0}$
  - linear in $\mathbf{a}$!

AR(4), impose F_1 resonance:

$$A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - a_3 z^{-3} - a_4 z^{-4}$$
$$Q(z) = 1 - c z^{-1} + d z^{-2}, \quad c = 2\rho \cos \omega_0, d = \rho^2$$
$$R(z) = 1 - r_1 z^{-1} - r_2 z^{-2}$$

$M \mid A \iff A(z) = Q(z)R(z)$. Match $z^{-1} ... z^{-4}$:

$$a_1 = c + r_1, \qquad a_2 = r_2 - c r_1 - d$$
$$a_3 = d r_1 - c r_2, \quad a_4 = d r_2$$

Eliminate r_1, r_2:

$$\underbrace{\begin{pmatrix} 1 & 0 & -1/d & -c/d^2 \\ 0 & 1 & c/d & c^2/d^2 - 1/d \end{pmatrix}}_{\mathbf{F}} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} + \underbrace{\begin{pmatrix} -c \\ d \end{pmatrix}}_{\mathbf{c}} = \mathbf{0}$$

Spectral feature prior

- Want a better prior $p(\mathbf{a})$
  - but not restrictive
- Idea:
  - impose $M \mid A$ on *average*
  - $\mathbb{E}_{\mathbf{a}}[\mathbf{F}\mathbf{a} + \mathbf{c}] = \mathbf{0}$
- Closed form:

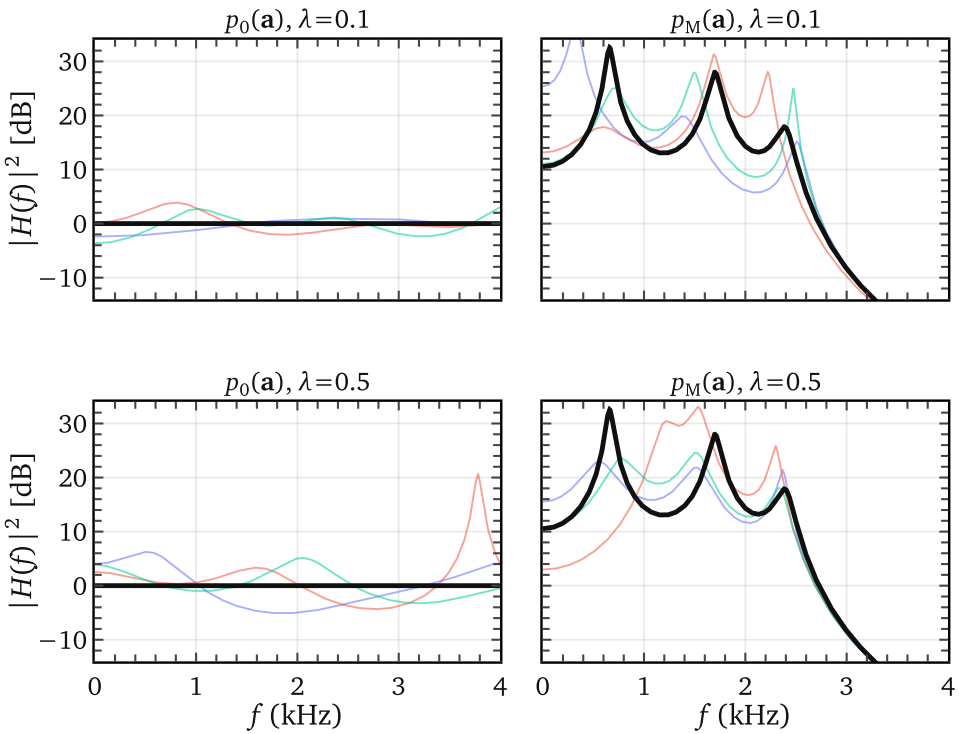
$p_0(\mathbf{a}) = \mathcal{N}(\mathbf{0}, \lambda\mathbf{I})$
uninformative prior

$\operatorname{argmin}_p D_{\mathrm{KL}}(p \parallel p_0)$

$\mathbb{E}_{\mathbf{a}}[\mathbf{F}\mathbf{a} + \mathbf{c}] = \mathbf{0}$

$p_M(\mathbf{a}) = \mathcal{N}(\boldsymbol{\mu}_M, \lambda\mathbf{I})$
quasi-informative prior

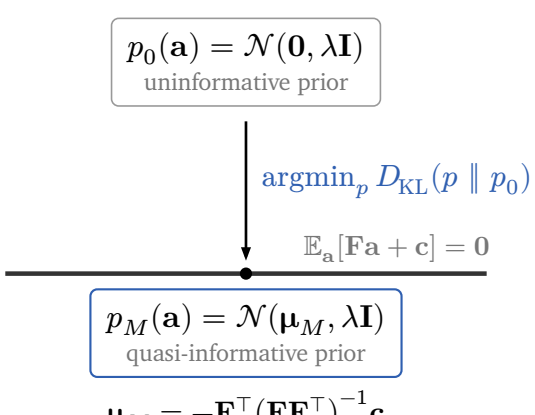
$\boldsymbol{\mu}_M = -\mathbf{F}^\top(\mathbf{F}\mathbf{F}^\top)^{-1}\mathbf{c}$

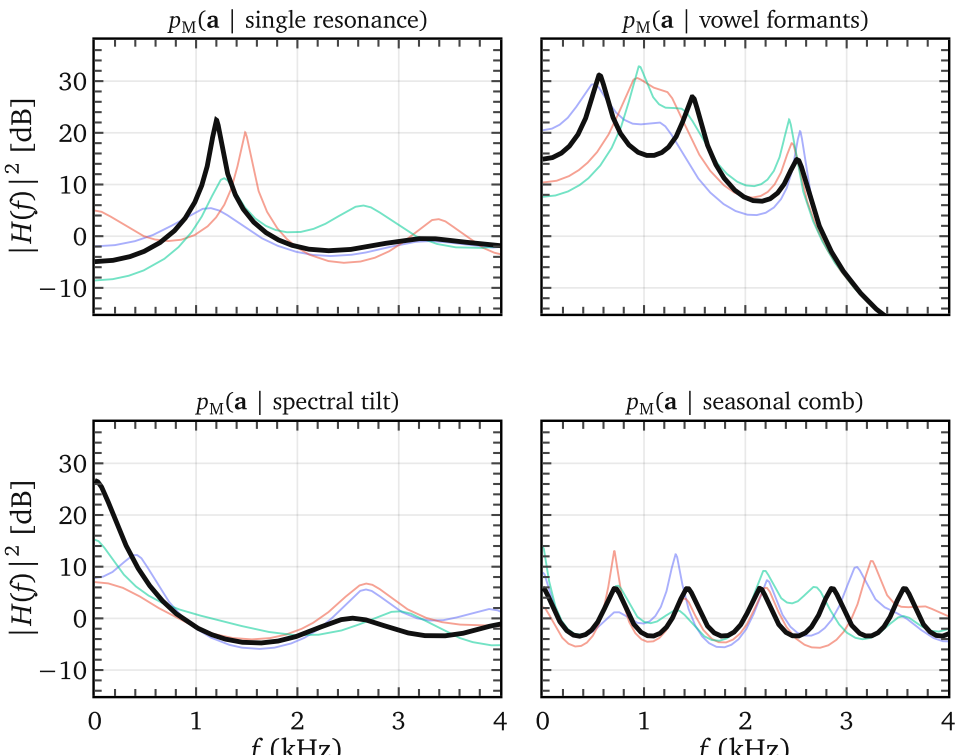


"Universal" AR(P) prior
[Cuevas et al. 2021; Sørbye & Rue 2017; Gibson 2018]

Spectral feature prior

- Want a better prior $p(\mathbf{a})$
  - but not restrictive
- Idea:
  - impose $M \mid A$ on *average*
  - $\mathbb{E}_{\mathbf{a}}[\mathbf{F}\mathbf{a} + \mathbf{c}] = \mathbf{0}$
- Closed form:





"Universal" AR(P) prior

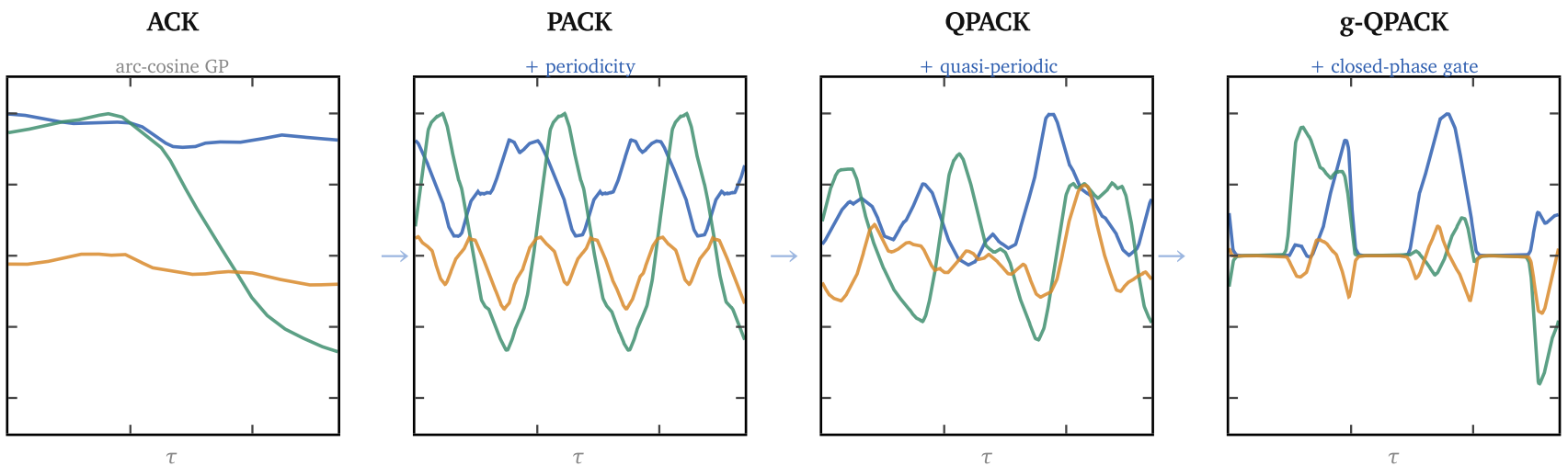
[Cuevas et al. 2021; Sørbye & Rue 2017; Gibson 2018]

4 Excitation prior

$$x(t) = e(t) * h(t)$$

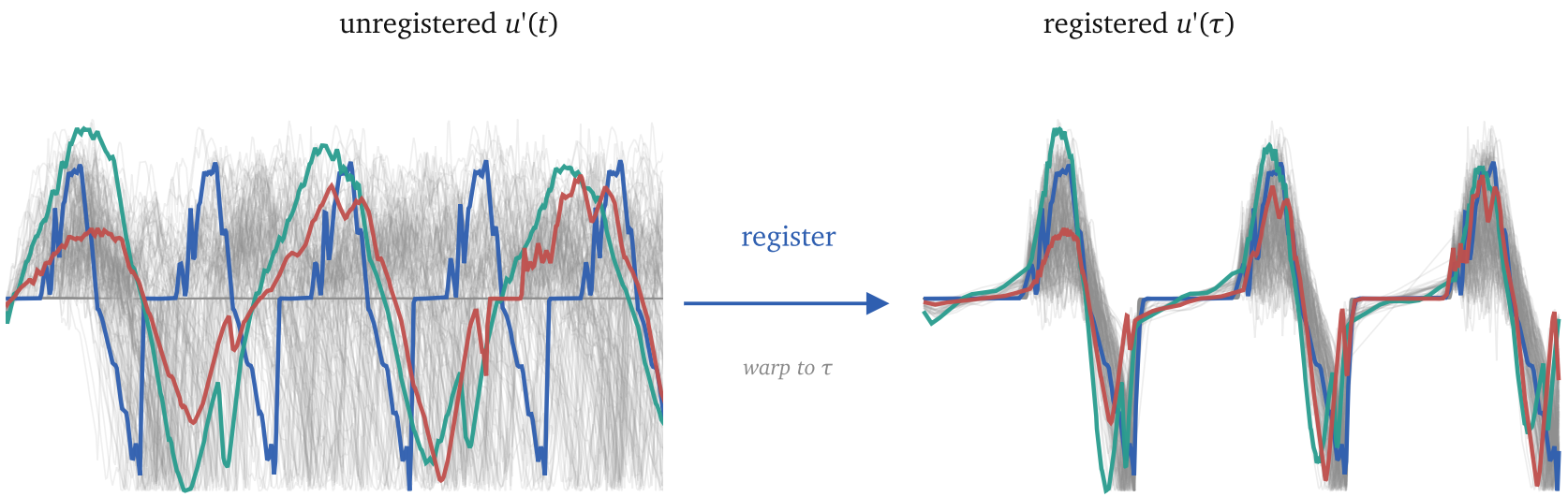
$$\begin{aligned} p(x, e, h) &= p(x \mid e, h)\, \boldsymbol{p(e)}\, p(h) \\ &= p(e, h \mid x)\, p(x) \end{aligned}$$

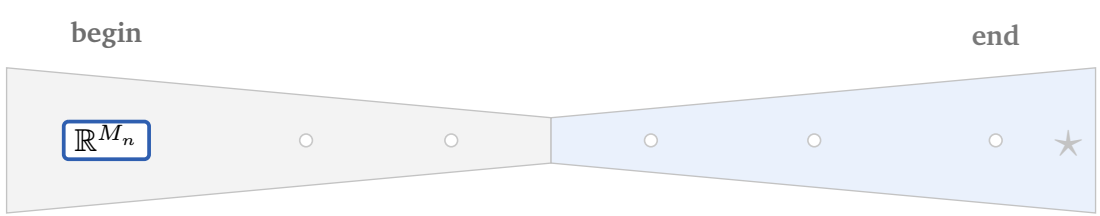
Kernel design



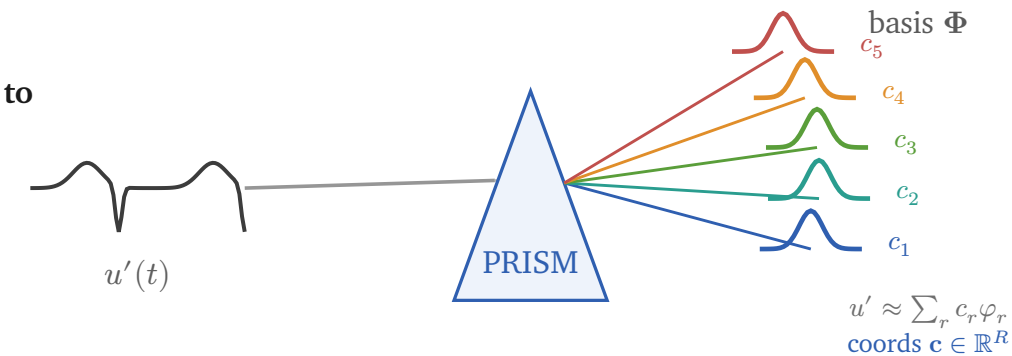
- Design a kernel with glottal flow features
  - GP limit of classical models [Cho & Saul 2009]
  - spectral tilt
  - quasi-periodicity
  - closed phase

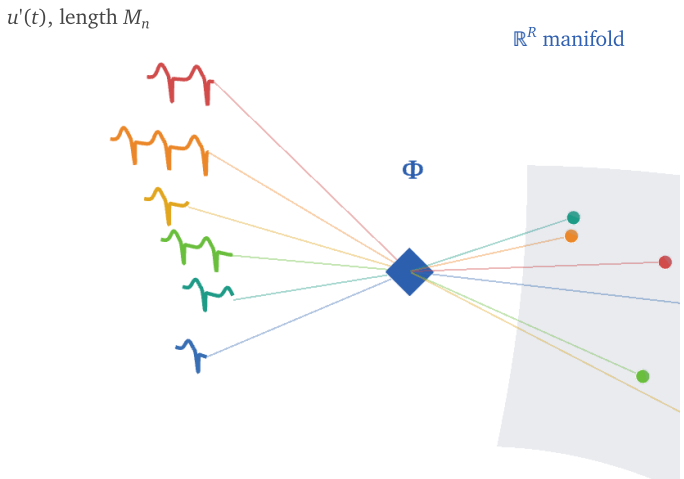
From corpus to kernel bank

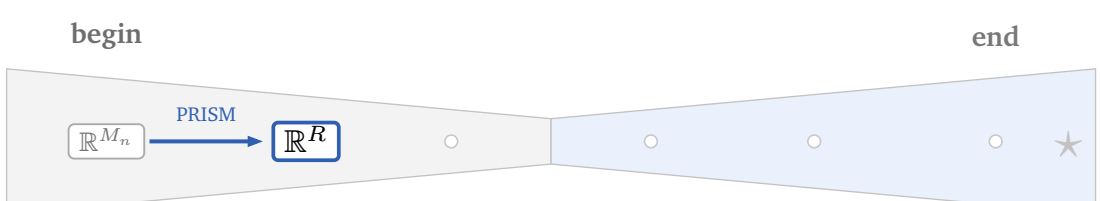




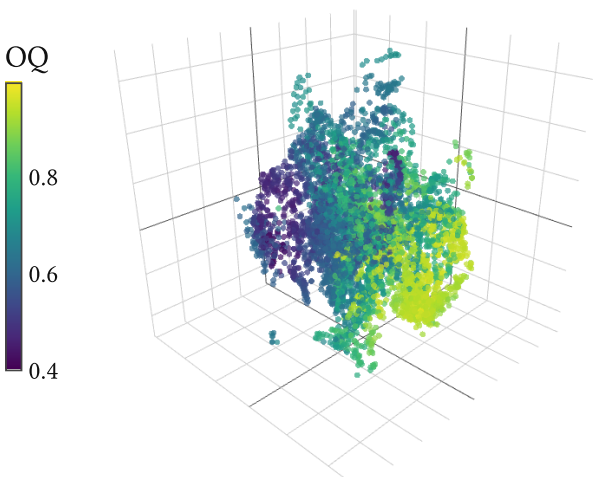
From corpus to kernel bank

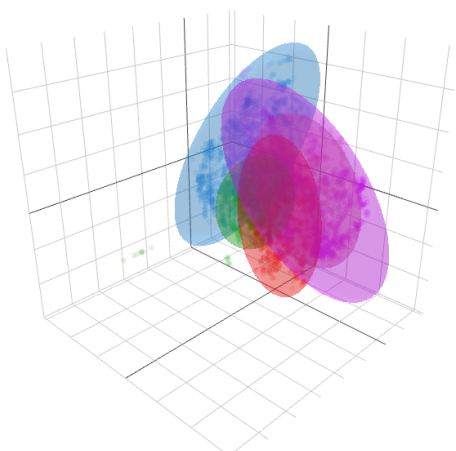


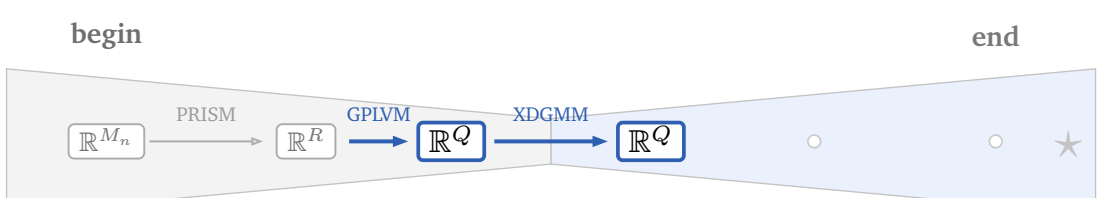




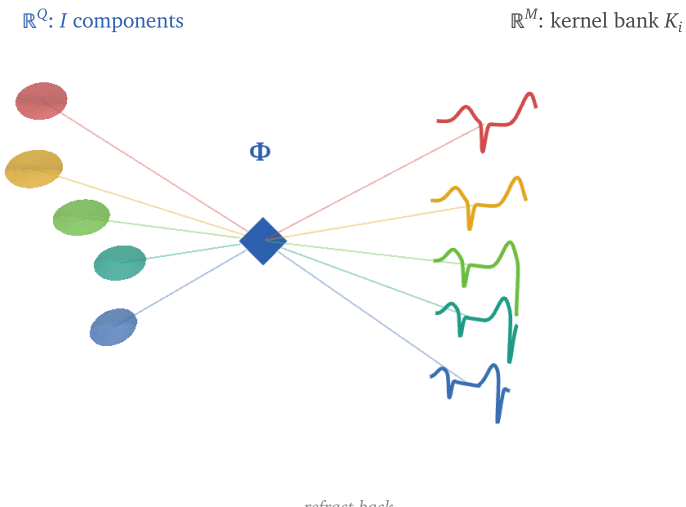
From corpus to kernel bank

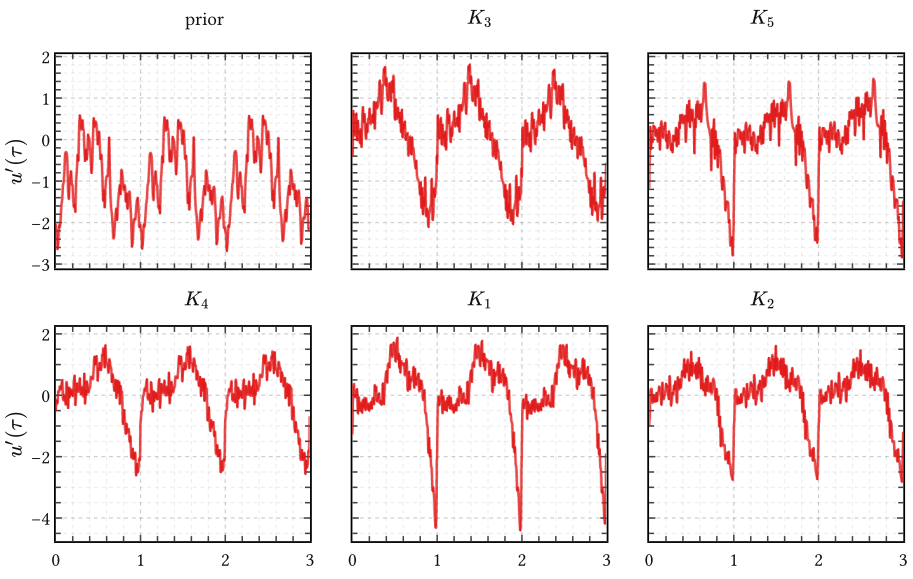


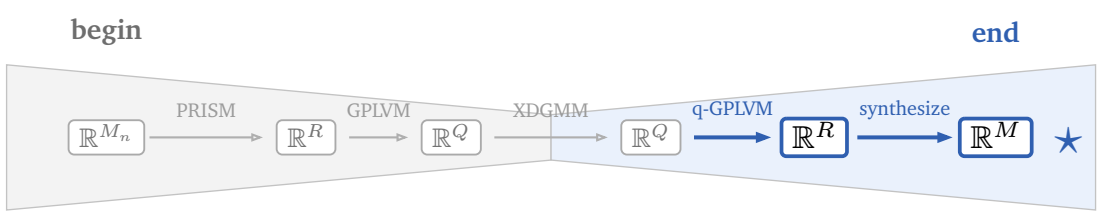




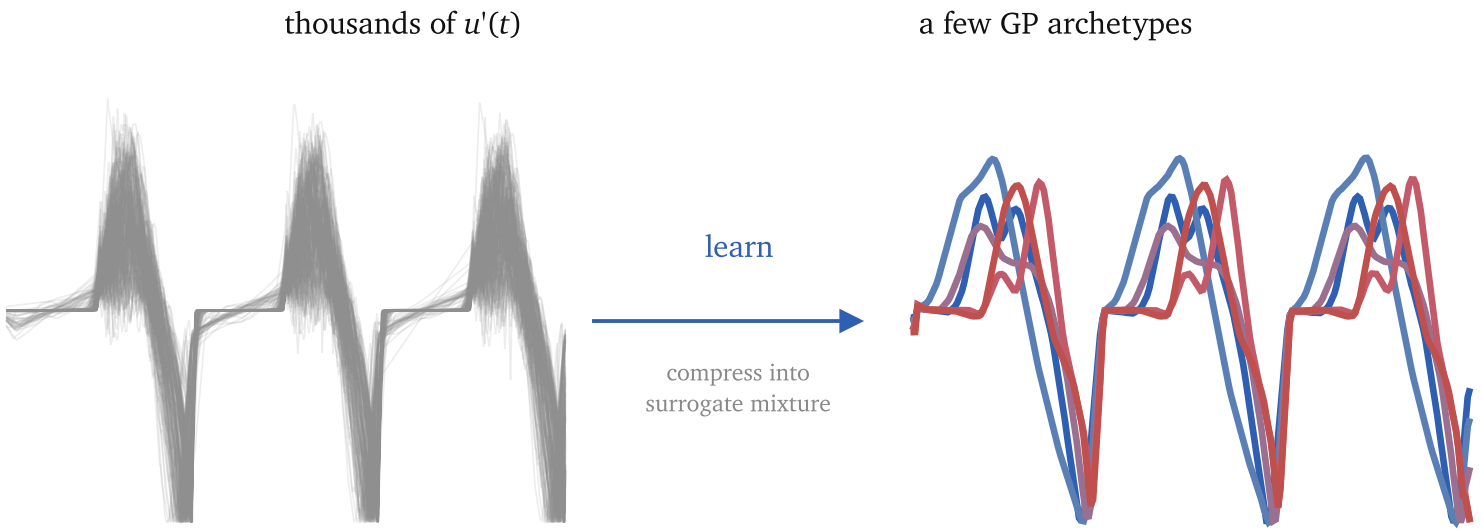
From corpus to kernel bank

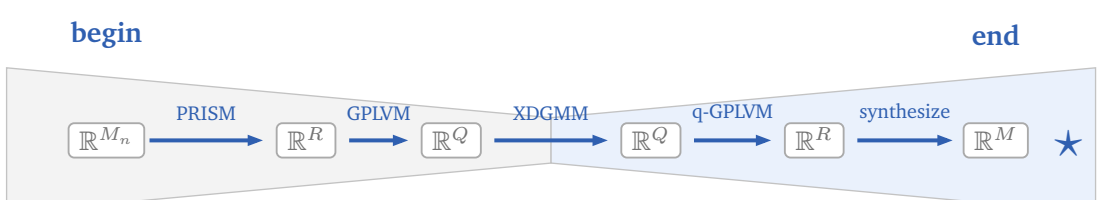






From corpus to kernel bank





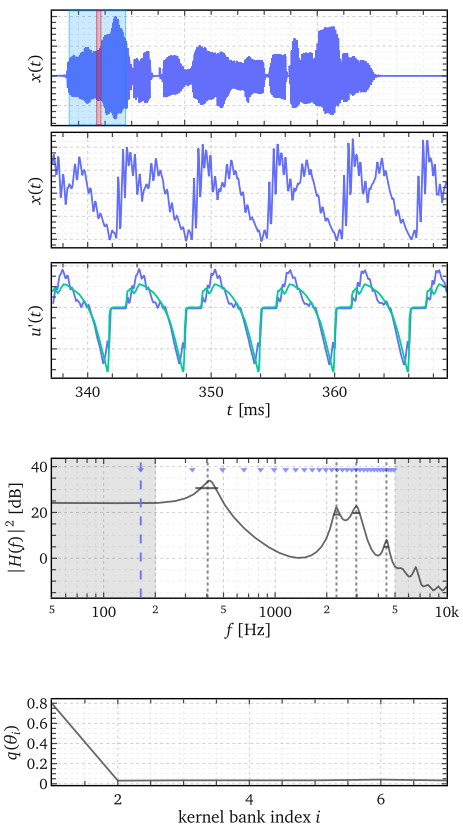
5 BNGIF

$$x(t) = e(t) * h(t)$$

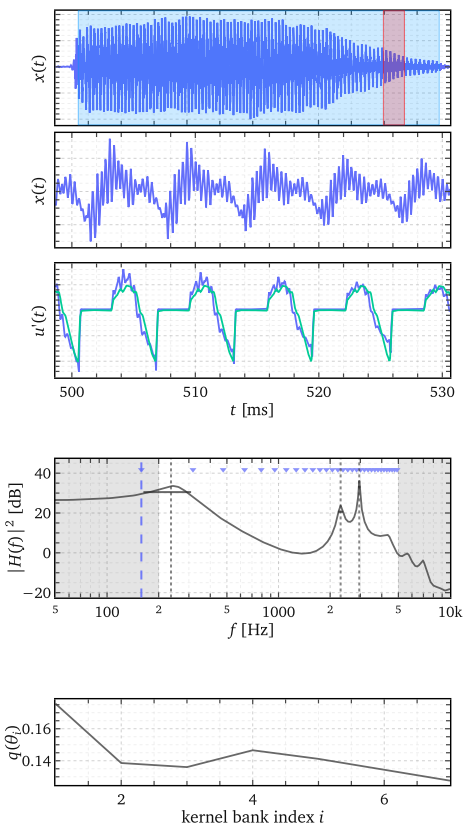
$$\begin{aligned} p(x, e, h) &= p(x \mid e, h)\, p(e)\, p(h) \\ &= \boldsymbol{p(e, h \mid x)}\, p(x) \end{aligned}$$

Application

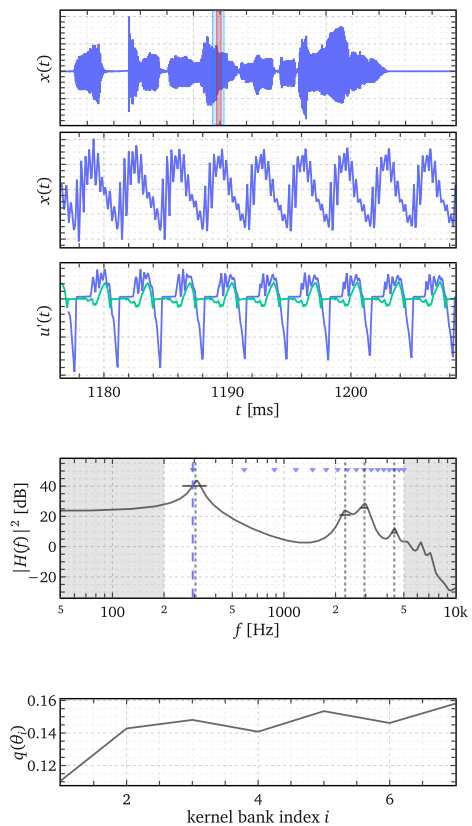
sparse



mixture

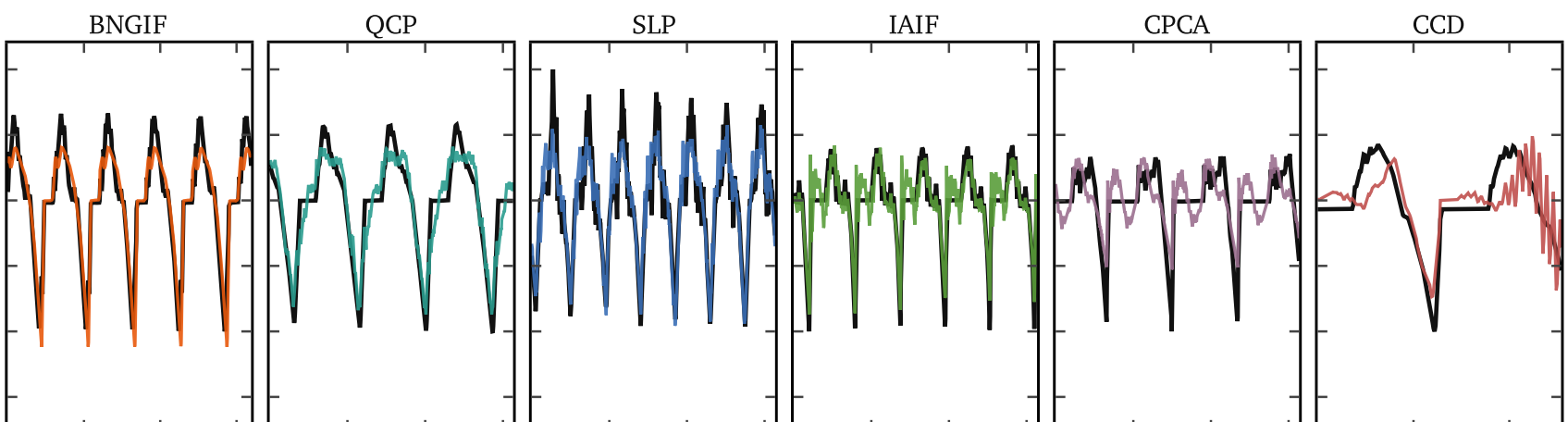


$F_1 \approx F_0$ (fails)



EGIFA benchmark

- Goal:
  - recover true $u(t)$
- Median performance for 6 methods



(ours) [Airaksinen et al. 2014] [Chien et al. 2017] [Alku 1992] [Wong et al. 1979] [Drugman et al. 2011]

EGIFA benchmark

- Goal:
  - recover true $u(t)$
- Paired Wilcoxon rank test

EGIFA / speech

	BNGIF	QCP	SLP	IAIF	CPCA	CCD
BNGIF	0.935	<.001	<.001	<.001	<.001	<.001
QCP	≈ 1	0.910	<.001	<.001	<.001	<.001
SLP	≈ 1	≈ 1	0.889	<.001	<.001	<.001
IAIF	≈ 1	≈ 1	≈ 1	0.886	<.001	<.001
CPCA	≈ 1	≈ 1	≈ 1	≈ 1	0.759	0.004
CCD	≈ 1	≈ 1	≈ 1	≈ 1	≈ 1	0.741

EGIFA / vowel

	QCP	SLP	BNGIF	IAIF	CPCA	CCD
QCP	0.957	<.001	0.37	<.001	<.001	<.001
SLP	≈ 1	0.946	0.97	<.001	<.001	<.001
BNGIF	0.64	0.03	0.935	<.001	<.001	<.001
IAIF	≈ 1	≈ 1	≈ 1	0.935	<.001	<.001
CPCA	≈ 1	≈ 1	≈ 1	≈ 1	0.844	<.001
CCD	≈ 1	≈ 1	≈ 1	≈ 1	≈ 1	0.820

  - legend:

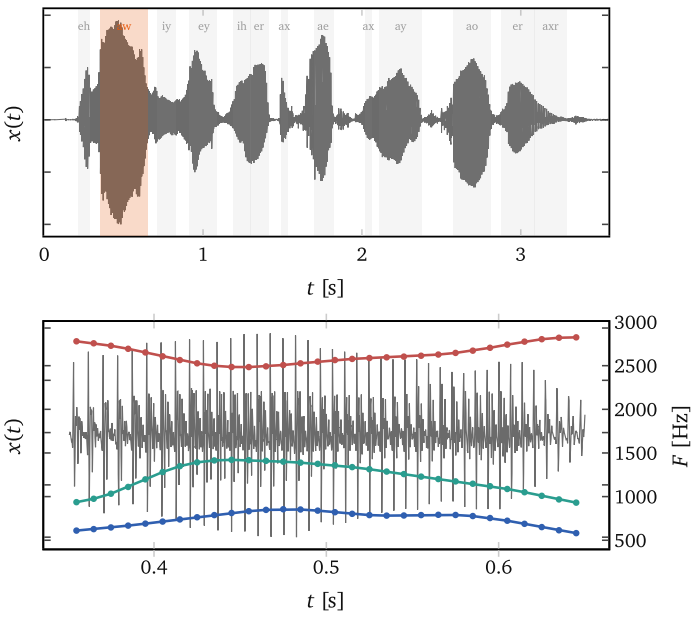
score	diagonal	off-diagonal	shaded
similarity to true $u'(t)$	median score	$p(\text{row} > \text{column})$, paired Wilcoxon	row better ($p < 0.05$)

VTRFormants benchmark

- Goal:
  - recover annotated F_1, F_2, F_3 tracks

- Average performance for 4 methods:

method	F_1	F_2	F_3	overall
BNGIF (informative)	**82**	**129**	**223**	**153**
BNGIF (uninformative)	89	134	225	157
KARMA [Mehta et al. 2012]	114	226	320	220
Praat [Boersma 2001]	185	254	303	247
WaveSurfer [Sjölander & Beskow 2000]	170	276	383	276



6 Conclusion

$$x(t) = e(t) * h(t)$$

$$\begin{aligned} p(x, e, h) &= p(x \mid e, h)\, p(e)\, p(h) \\ &= p(e, h \mid x)\, p(x) \end{aligned}$$

Summary

- Blind audio deconvolution
  - glottal inverse filtering
  - fundamental problem
  - grey-box model
- Principled Bayesian regularization
  - via priors
  - process information via D_{KL}
    [Jaynes 2003; Knuth & Skilling 2012]
  - uncertainty quantification
- And it works:
  - SOTA on EGIFA + VTRFormants
  - $>$ realtime (was 2.5 h / frame)
    [Auvinen et al. 2014]

…mad party

Summary

- General tools developed
  - Spectral feature prior
  - PRISM
  - q-GPLVM
  - *learn* GPs from corpora…
  - … and *sparsify* them
- Important applications:
  - fundamental speech science
  - clinical medicine [Alku et al. 2026]
  - forensic [Becker et al. 2008]

…mad party

Future work

- Limitations
  - $F_1 \approx F_0$ stays hard
  - needs closure event information
  - pipeline: one by one, not unified
  - frames are independent
- Future work
  - dependent frames
  - end-to-end pipeline
  - downstream validation for clinical medicine
  - other simulators [Schoder et al. 2024]
  - coupled AR-GP [Yoshii & Goto 2012]

which way?

Thank you for listening!

References

Airaksinen, M., Raitio, T., Story, B., & Alku, P. (2014). Quasi Closed Phase Glottal Inverse Filtering Analysis With Weighted Linear Prediction. *IEEE/ACM Transactions on Audio, Speech, And Language Processing*, *22*(3), 596–607. https://doi.org/10.1109/TASLP.2013.2294585

Alku, P. (1992). *Glottal Wave Analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering*. *11*.

Alku, P., Hernández-Villoria, R., & Das, S. (2026). Glottal Inverse Filtering and Its Application in the Automatic Classification of Diseases from Speech. In M. Pedersen, N. Hassan Nashaat, V. Camesasca, R. Hernández Villoria, & S. Das (Eds.), *Voice-Related Biomarkers: Voice-Related Biomarkers* (pp. 53–71). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-03134-1_5

Atal, B. S., & Hanauer, S. L. (1971). Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. *The Journal of the Acoustical Society of America*, *50*(2B), 637–655.

Auvinen, H., Raitio, T., Airaksinen, M., Siltanen, S., Story, B. H., & Alku, P. (2014). Automatic Glottal Inverse Filtering with the Markov Chain Monte Carlo Method. *Computer Speech & Language*, *28*(5), 1139–1155. https://doi.org/10.1016/j.csl.2013.09.004

Becker, T., Jessen, M., & Grigoras, C. (2008,). Forensic Speaker Verification Using Formant Features and Gaussian Mixture Models. *Ninth Annual Conference of the International Speech Communication Association*.

Boersma, P. (2001). *Praat, a System for Doing Phonetics by Computer*. *5*(9), 341–345.

Chien, Y.-R., Mehta, D. D., Guenason, J., Zanartu, M., & Quatieri, T. F. (2017). Evaluation of Glottal Inverse Filtering Algorithms Using a Physiologically Based Articulatory Speech Synthesizer. *IEEE/ACM Transactions on Audio, Speech, And Language Processing*, *25*(8), 1718–1730. https://doi.org/10.1109/TASLP.2017.2714839

Cho, Y., & Saul, L. (2009). Kernel Methods for Deep Learning. *Advances in Neural Information Processing Systems*, *22*.

Cuevas, A., López, S., Mandic, D., & Tobar, F. (2021). Bayesian Autoregressive Spectral Estimation. *2021 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 1–7. https://doi.org/10.1109/LA-CCI48322.2021.9769834

Drugman, T., Bozkurt, B., & Dutoit, T. (2011). Causal–Anticausal Decomposition of Speech Using Complex Cepstrum for Glottal Source Estimation. *Speech Communication*, *53*(6), 855–866. https://doi.org/10.1016/j.specom.2011.02.004

Gibson, J. (2018). Entropy Power, Autoregressive Models, and Mutual Information. *Entropy*, *20*(10), 750. https://doi.org/10.3390/e20100750

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jopling, R. L. (2009). *Voice Initiation and Voice Offset Patterns in Normal Females: Investigated by High Speed Digital Imaging* [Doctoral dissertation].

Knuth, K. H., & Skilling, J. (2012). Foundations of Inference. *Axioms*, *1*(1), 38–73. https://doi.org/10.3390/axioms1010038

Mehta, D. D., Rudoy, D., & Wolfe, P. J. (2012). Kalman-Based Autoregressive Moving Average Modeling and Inference for Formant and Antiformant Tracking. *The Journal of the Acoustical Society of America*, *132*(3), 1732–1746. https://doi.org/10.1121/1.4739462

Rabiner, L. R., Juang, B.-H., & Rutledge, J. C. (1993). *Fundamentals of Speech Recognition* (Vol. 14). PTR Prentice Hall Englewood Cliffs.

Schoder, S., Falk, S., Wurzinger, A., Lodermeyer, A., Becker, S., & Kniesburges, S. (2024). A Benchmark Case for Aeroacoustic Simulations Involving Fluid-Structure-Acoustic Interaction Transferred from the Process of Human Phonation. *Acta Acustica*, *8*, 13. https://doi.org/10.1051/aacus/2024005

Sjölander, K., & Beskow, J. (2000). Wavesurfer - an Open Source Speech Tool. *6th International Conference on Spoken Language Processing (ICSLP 2000)*, vol. 4, 464–467–0. https://doi.org/10.21437/ICSLP.2000-849

Sørbye, S. H., & Rue, H. (2017). Penalised Complexity Priors for Stationary Autoregressive Processes. *Journal of Time Series Analysis*, *38*(6), 923–935. https://doi.org/10.1111/jtsa.12242

Wong, D., Markel, J., & Gray, A. (1979). Least Squares Glottal Inverse Filtering from the Acoustic Speech Waveform. *IEEE Transactions on Acoustics, Speech, And Signal Processing*, *27*(4), 350–355. https://doi.org/10.1109/TASSP.1979.1163260

Yoshii, K., & Goto, M. (2012b). *INFINITE COMPOSITE AUTOREGRESSIVE MODELS FOR MUSIC SIGNAL ANALYSIS*.

Yoshii, K., & Goto, M. (2013a). Infinite Kernel Linear Prediction for Joint Estimation of Spectral Envelope and Fundamental Frequency. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 463–467. https://doi.org/10.1109/ICASSP.2013.6637690