SCIENCES & BIOENGINEERING SCIENCES

The Research Group
Artificial Intelligence Lab

has the honor to invite you to the public defence of the PhD thesis of

Marnix Van Soom

to obtain the degree of Doctor of Sciences

Title of the PhD thesis:

Bayesian Nonparametric Glottal Inverse Filtering

Supervisor:
Prof. dr. Bart de Boer (VUB)
Co-supervisor:
Dr. Ivan De Boi (UAntwerpen)

The defence will take place on
Friday, September 18, 2026 at 4 p.m.

VUB Etterbeek campus, Pleinlaan 2, Elsene, in promotiezaal D.2.01

The defence can be followed through a live stream: Microsoft Teams link

Members of the jury

Prof. dr. Elisa Gonzalez Boix (VUB, chair)
Prof. dr. Ann Nowé (VUB)
Prof. dr. ir. John Lataire (VUB)
Prof. dr. Rudi Penne (UAntwerpen)
Prof. dr. Inneke van Nieuwenhuyse (UHasselt)
Dr. Henry Moss (Lancaster University, UK)

Curriculum vitae

I'm a researcher and developer working across Bayesian machine learning, software, and media art. I come from a background in programming, physics and applied mathematics. My research has appeared in *Scientific Reports* and *Entropy*. I have worked on signal processing, entropic inference, network theory and language communication.

Abstract of the PhD research

Speech can be described locally as the convolution of a glottal source wave and a vocal tract filter. We can measure speech directly, but not the source and filter individually. Recovering them from speech alone is a severely underdetermined inverse problem called glottal inverse filtering (GIF), open for over fifty years. GIF is a special case of blind audio deconvolution. Next to its relevance for fundamental speech science, GIF matters for early diagnosis of neurodegenerative diseases such as Parkinson's, because these interfere with speech production long before other motor signs appear.

To handle this underdetermination, we recast GIF as Bayesian inference and call our approach Bayesian nonparametric glottal inverse filtering (BNGIF). We learn a Gaussian process prior over glottal waveforms from a large corpus of physics-based simulations, embed the data in a small latent space, and obtain a compact mixture of surrogate models that serves as the source prior. The vocal tract filter is regularized by a new information-theoretic framework for encoding spectral constraints on the AR coefficients.
Both sides are inferred jointly by a fast variational method that runs at 5x realtime on a single commodity GPU.

We validate on two important tasks: glottal waveform recovery and formant estimation. On glottal waveform recovery, BNGIF achieves the strongest speech scores among state of the art and is on par with the best classical method on steady-state vowels. On formant estimation, it reduces overall RMSE by 30% relative to the best classical tracker. Both results come from the same posterior: better source modeling yields better filter estimates, and vice versa.

These scores are encouraging, but the thesis does not so much "solve" GIF as push it to what seems a more natural setting for it. That is, our main contribution is a coherent attempt to cast GIF as Bayesian inference without suffering a practical cost in terms of usage: inference is stable and still fast enough such that BNGIF could be part of daily practice. The Bayesian handle we develop below opens the door to transparent model criticism, data-driven learning, principled regularization, and calibrated uncertainty – properties that the current state of the art, for all its practicality and ingenuity, does not yet offer.