# A SNACK IMPLEMENTATION AND TcI/Tk INTERFACE TO THE FUNDAMENTAL FREQUENCY VARIATION SPECTRUM ALGORITHM

Kornel Laskowski<sup>1</sup> and Jens Edlund<sup>2</sup>

<sup>1</sup> Carnegie Mellon University, Pittsburgh PA, USA

<sup>2</sup> KTH Speech, Music and Hearing, Stockholm, Sweden

# Goal

In real-time speech processing applications, **intonation** may be useful as early as possible, even prior to lexical segmentation.

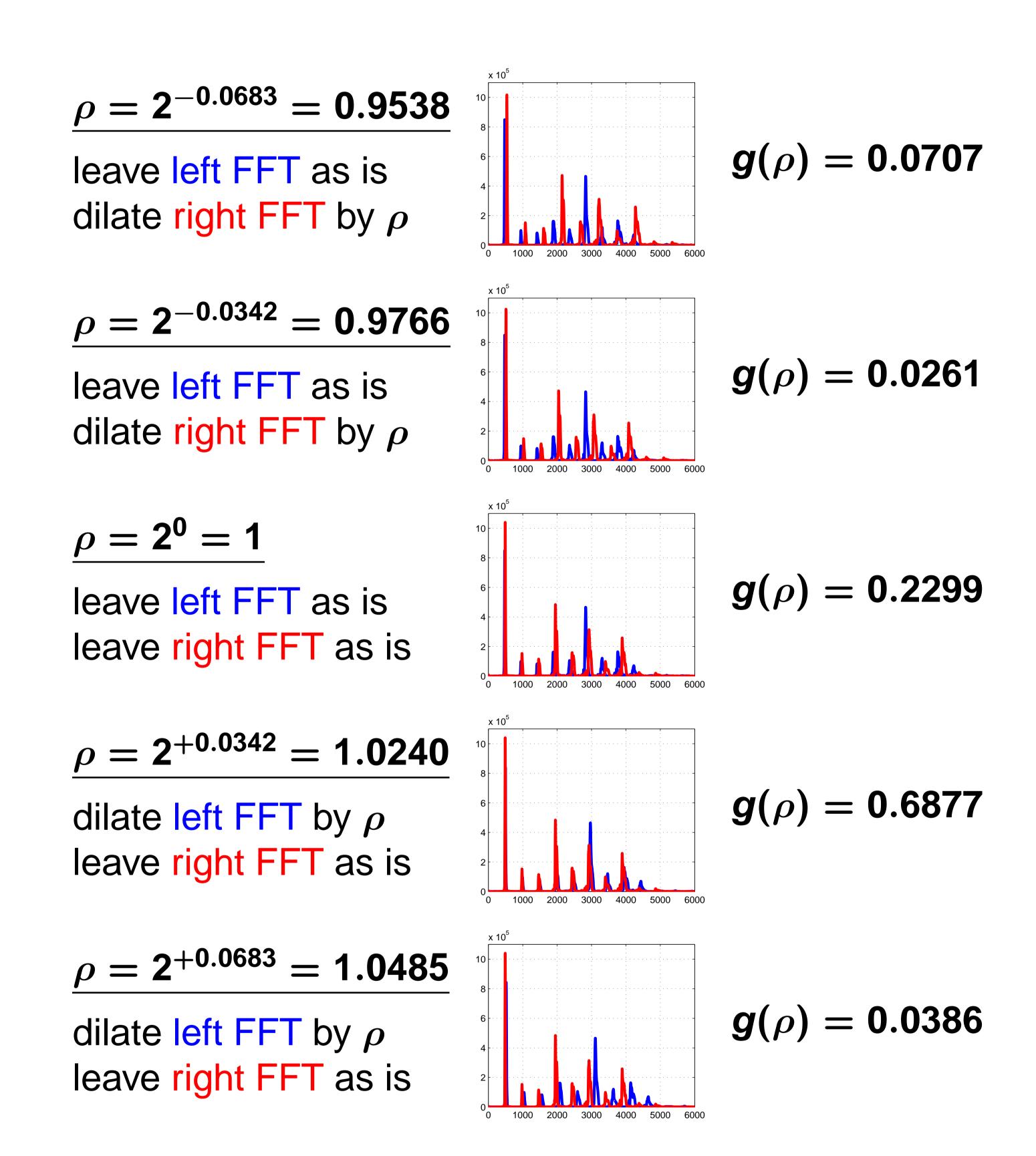
#### However:

- 1. Intonation is typically estimated by computing a pitch trajectory.
- 2. Pitch estimation is performed only after speech segmentation (and, often, recognition).
- 3. Therefore, intonation is currently unavailable to early speech processing components.

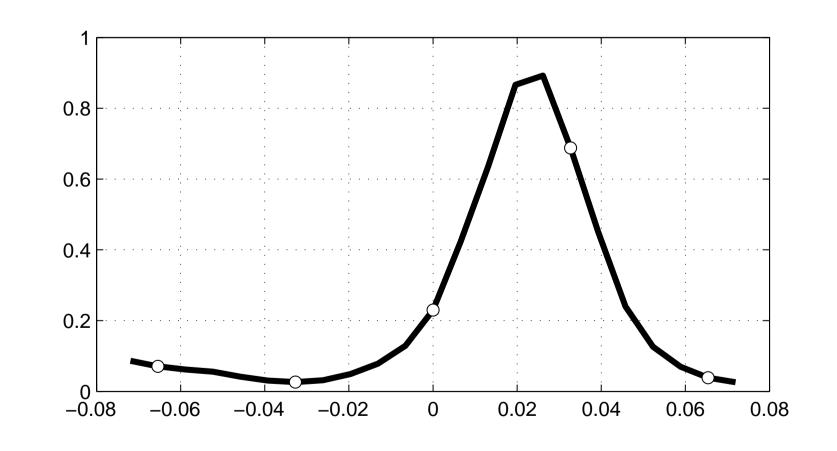
This work describes a new, freely available implementation of an alternative to pitch estimation: the computation of the fundamental frequency variation (FFV) spectrum.

# What is the FFV Spectrum?

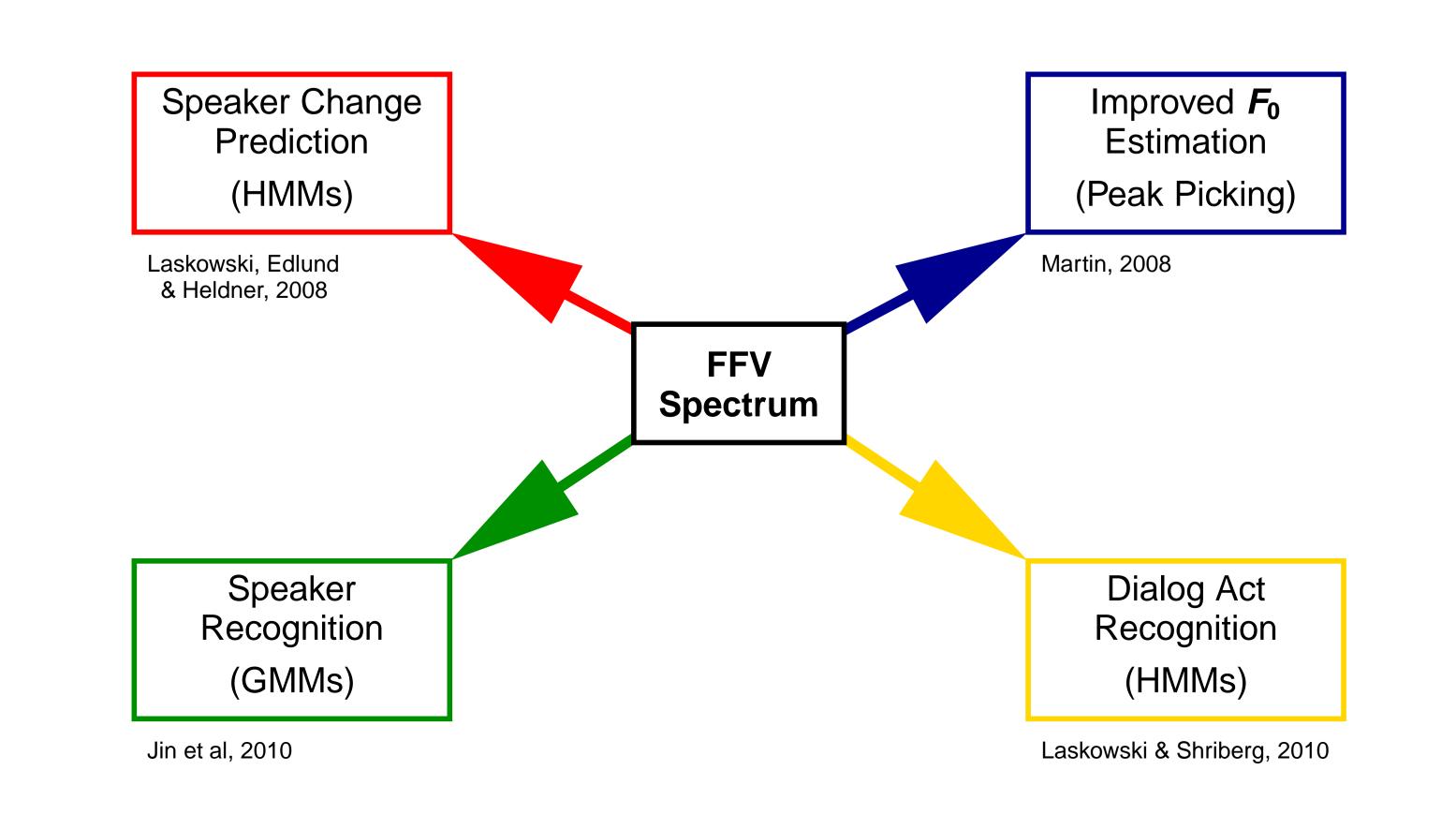
- the dot product g of two adjacent FFT spectra
- ullet after frequency dilation by a factor ho
- ullet for a range of ho values



The FFV spectrum is a plot of  $g(\rho)$  as a function of  $\log_2 \rho$ :



# **Explored Use of FFV, To Date**



# **Conclusions & Impact**

- 1. Peaks in the FFV spectrum provide visually identical information to the first derivative of a pitch trajectory.
- 2. Computation of the FFV spectrum is fully and flexibly implemented in Snack, a popular, freely available speech processing toolkit.
- 3. Accessible with a single line of code.
- 4. FFV spectrum can be explored by anyone, with almost no learning curve.
- 5. Intonation is available to front-end components of speech processing systems, much earlier than previously considered.

#### What is Snack?

The Snack Sound Toolkit, developed by Kåre Sjölander, allows the creation of multi-platform audio applications with just a few lines of code.

- freely available
- easily extendible
- contains primitives for visualization
- bindings for Tcl/Tk, Python, and Ruby

## The New Tcl Interface to FFV in Snack

Just like pitch estimation in Snack, for a sound object s:

s pitch -method ESPS
-minpitch 60 -maxpitch 1000
-framelength 0.008 -windowlength 0.0075

## FFV is now available with a single Tcl command:

s ffv -tFra 0.008 -filterbank NONE -tSep 0.112 -tInt 0.088 -tExt 0.072

## Visual Comparison of Pitch Tracker and FFV Output

- compare Snack FFV output with the first-order difference of the Snack ESPS pitch trajectory
- use singing voice as a novel example
  - unambiguous, language-independent expectations
  - no unvoiced occlusions

