Bert Remijsen, "Tonal alignment is contrastive in falling contours in Dinka" *Language* 2013:

This study investigates a contrast in tonal alignment, involving falling contours in Dinka. Such a contrast, first reported in Andersen (1987), calls into question the assumption that tonal alignment cannot distinguish contour tone patterns of the same shape within the syllable domain. […] The primary correlate of the contrast is indeed tonal alignment: the early-aligned fall sets in during the onset or early in the vowel; the late-aligned peak sets in well into the vowel. […] [I]t is argued that a representation with reference to mora […] is less attractive: this representation is not sufficiently restrictive. Instead, the contrast is represented using a binary feature [±late-aligned].

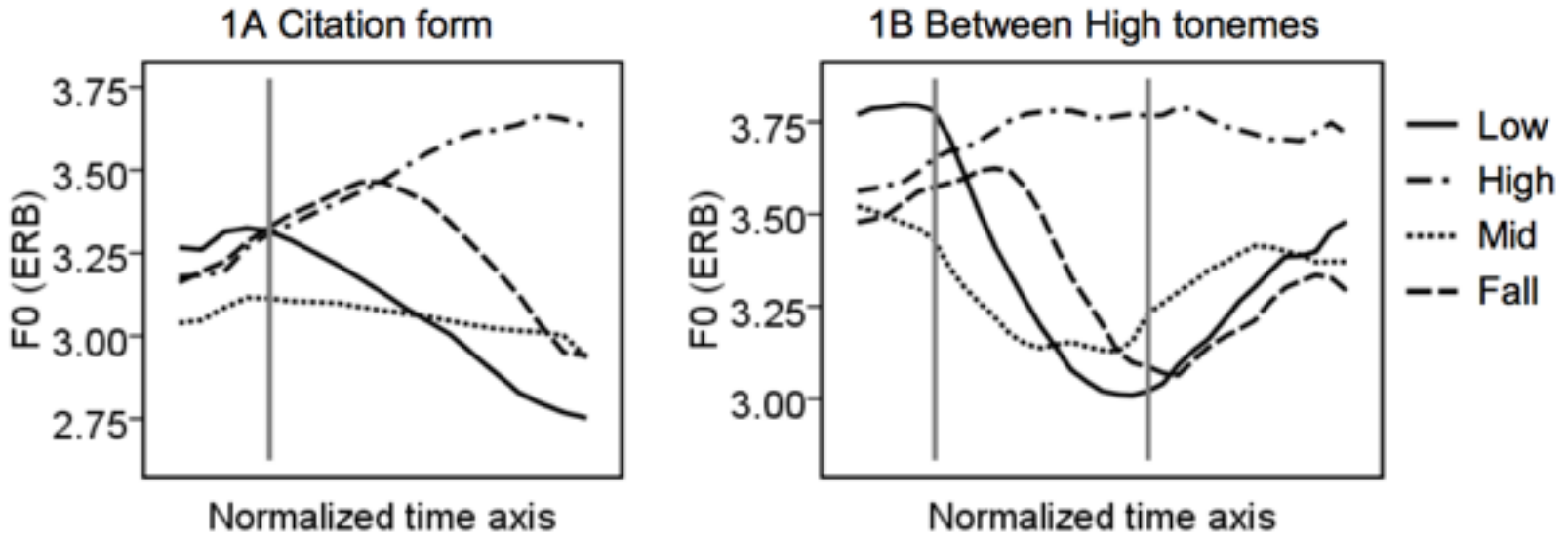
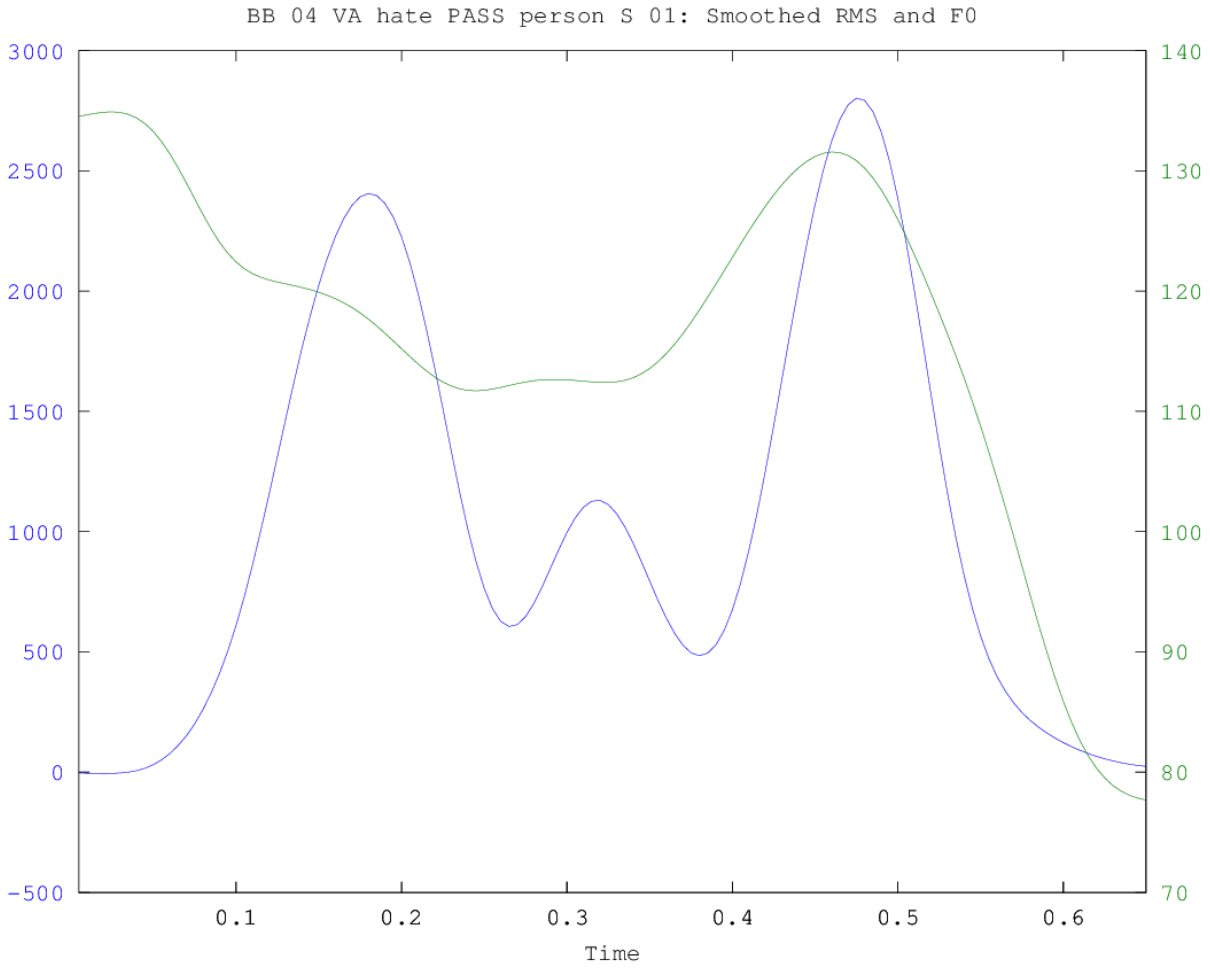
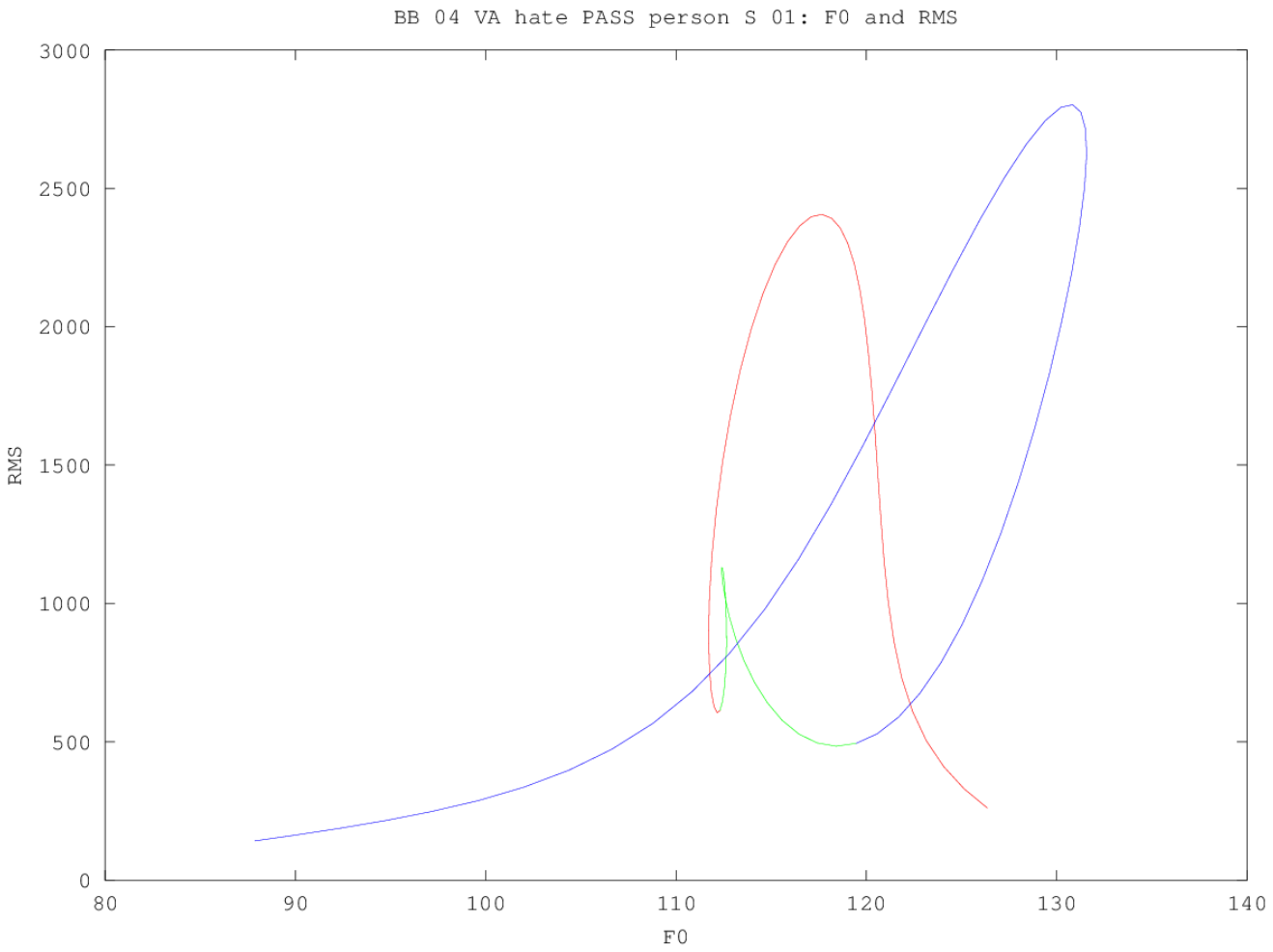


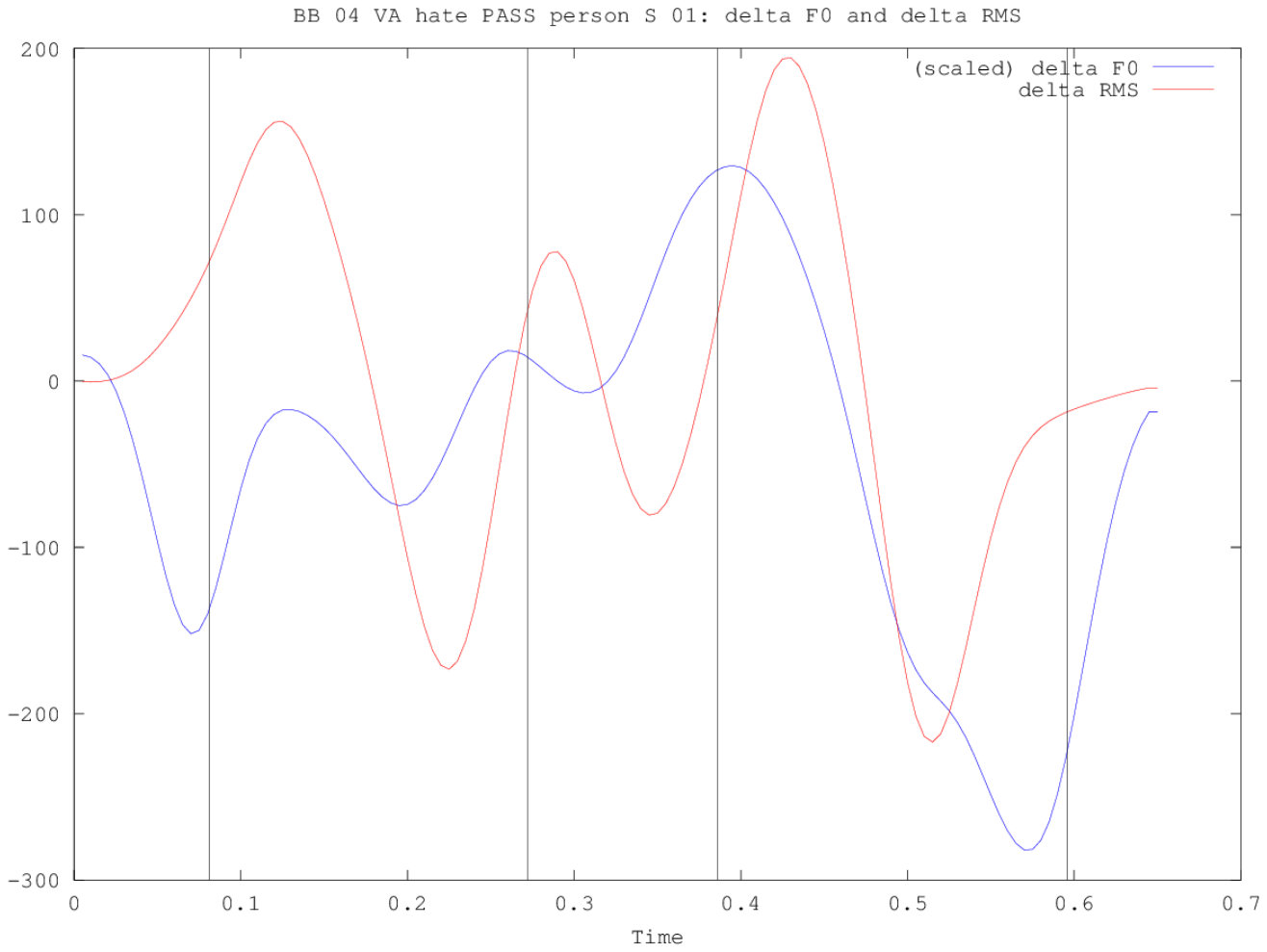
Figure 1. Averaged interpolated f0 traces on a normalized time axis, showing the realization of the four tonemes – Low, High, Fall, and Mid – in /nòoon/ 'grass\S', /lɔ̰ɔ̰ɔm/ 'rib\S', /ŋâaap/ 'sycamore\S', and /lāaac/ 'urine\U', respectively. Each trace is averaged across realizations by three speakers of Bor North. Panel 1A shows citation forms. Panel 1B shows the same target words in a frame sentence, between High tonemes. The vertical line in 1A and the one on the left in 1B both mark the start of the vowel of the target word. The second vertical line in 1B marks the end of the target word.


BB 04 VA hate PASS person S 01: Smoothed RMS and F0
3000
2500
2000
1500
1000
500
0
-500
140
130
120
110
100
90
80
70
0.1
0.2
0.3
0.4
0.5
0.6
Time

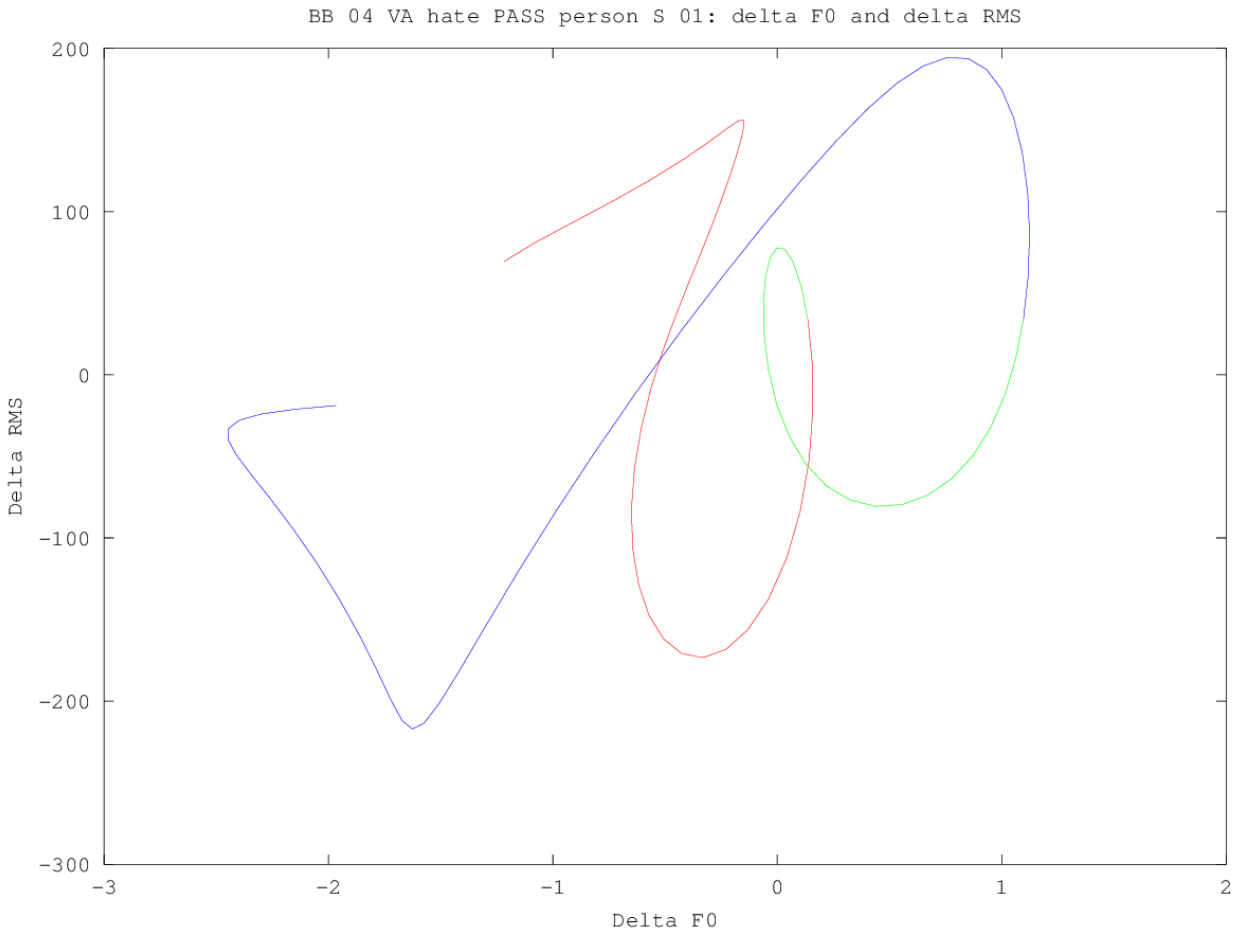
Red = 1st syllable, Green = 2nd syllable, Blue = 3rd syllable:



Vertical black lines: Start of vowels 1, 2 & 3, end of vowel 3:



Red = 1st syllable, Green = 2nd syllable, Blue = 3rd syllable:



Hypothesis:

A vector-valued time function
of (smoothed) F0 & (spectrally-weighted) amplitude

…will provide more-or-less complete information
about the nature and syllabic alignment of pitch contours –
without (and better than) phonetic segmentation.

Further hypothesis:

Functional principal components analysis of such time functions
will yield a low-dimensional characterization of prosodic regions,
which would differentiate tonemic categories
while also preserving relevant phonetic variation.

Because Bert published his data,

it's trivial to make a small test of Hypothesis 1:

e.g. 105 examples of a Dinka minimal pair

(a) *allocate_NTSXXX3SG** (N=45) [early aligned]
(b) *provoke_PASS** (N=60) [late aligned]

Five time differences (Blue=phonetics; Red=RMS):

F0 peak relative to onset start: F0oS
F0 peak relative to onset end: F0oE
F0 peak relative to nucleus end: F0nE
F0 peak relative to amplitude min: F0amin
F0 peak relative to amplitude max: F0amax

For the early-aligned words:

	aF0oS	aF0oE	aF0nE	aF0amin	aF0amax
mean	0.092	0.008	-0.096	0.053	-0.030
stdev	0.017	0.013	0.016	0.014	0.013
stderr	0.003	0.002	0.002	0.002	0.002

For the late-aligned words:

	bF0oS	bF0oE	bF0nE	bF0amin	bF0amax
mean	0.118	0.037	-0.071	0.084	-0.008
stdev	0.019	0.019	0.023	0.018	0.023
stderr	0.002	0.002	0.003	0.002	0.003

Correlation matrix for early-aligned words:

	aF0oS	aF0oE	aF0nE	aF0amin	aF0amax
aF0oS	1.00	0.51	0.22	0.77	0.32
aF0oE	0.51	1.00	0.66	0.89	0.80
aF0nE	0.22	0.66	1.00	0.53	0.66
aF0amin	0.77	0.89	0.53	1.00	0.70
aF0amax	0.32	0.80	0.66	0.70	1.00

Correlation matrix for the late-aligned words:

	bF0oS	bF0oE	bF0nE	bF0amin	bF0amax
bF0oS	1.00	0.72	0.54	0.87	0.66
bF0oE	0.72	1.00	0.81	0.90	0.91
bF0nE	0.54	0.81	1.00	0.70	0.87
bF0amin	0.87	0.90	0.70	1.00	0.82
bF0amax	0.66	0.91	0.87	0.82	1.00

Difference of mean differences – early- minus late-aligned:

F0oS	F0oE	F0nE	F0amin	F0amax
-0.026	-0.029	-0.025	-0.031	-0.022

Pooled standard deviations:

F0oS	F0oE	F0nE	F0amin	F0amax
0.136	0.128	0.141	0.128	0.137

And the resulting effect sizes (”Cohen’s d”):

F0oS	F0oE	F0nE	**F0amin**	F0amax
0.189	0.225	0.176	**0.242**	0.161

F0-peak- time minus preceding-amplitude-minimum wins!!!

Bert's reaction:

What an amazing result. I was impressed that the standard deviations for the measures of f0 peak relative to (i) amplitude minimum vs. (ii) to the end of the onset consonant are only a millisecond apart. Over the two sets (early aligned vs. late-aligned), the standard deviations must be smaller for the measures relative to amplitude mimum, given that the inferential test result is higher.

So, in retrospect, I could have saved myself several weeks. I find that unbelievable. The study of alignment is an area where most researchers would not consider using automated extraction.

My reaction:

Saving human labor is good. But we're close to being able to automate traditional phonetic transcription with human-like accuracy, anyhow.

A more important possibility is that F0 timing should be investigated relative to amplitude (e.g. sonority) contours, not segmental "magic moments".