Automatic Prosodic Annotation with Momel and Intsint

OTIM Workshop

Daniel Hirst

Laboratoire Parole et Langage, CNRS and Université de Provence
daniel.hirst@lpl-aix.fr

2011 May 23

Automatic Prosodic Annotation

Input sound file

Automatic Prosodic Annotation

Input	sound file
Detect F0	Optimise max and min f_0

Automatic Prosodic Annotation

Input	sound file
Detect F0	Optimise max and min f_0
Momel	Model f_0 as a quadratic spline function defined by a sequence of target points

Automatic Prosodic Annotation

Input	sound file
Detect F0	Optimise max and min f_0
Momel	Model f_0 as a quadratic spline function defined by a sequence of target points
INTSINT	Code target points as {T,M,B,H,S,L,U,D} (optimised)

Automatic Prosodic Annotation

Input	sound file
Detect F0	Optimise max and min f_0
Momel	Model f_0 as a quadratic spline function defined by a sequence of target points
INTSINT	Code target points as {T,M,B,H,S,L,U,D} (optimised)
Output	TextGrid with INTSINT targets aligned with signal.

Detect f_0

- first pass - detect with default parameters min = 60, max = 750

Detect f_0

- first pass - detect with default parameters min = 60, max = 750
- calculate first and third quartiles of f_0

Detect f_0

- first pass - detect with default parameters min = 60, max = 750
- calculate first and third quartiles of f_0
- min f_0 = q1 * 0.75

Detect f_0

- first pass - detect with default parameters min = 60, max = 750
- calculate first and third quartiles of f_0
- min f_0 = q1 * 0.75
- max f_0 = q3 * 2.5

Detect f_0

- first pass - detect with default parameters min = 60, max = 750
- calculate first and third quartiles of f_0
- min f_0 = q1 * 0.75
- max f_0 = q3 * 2.5
- second pass - detect with these min and max

Momel

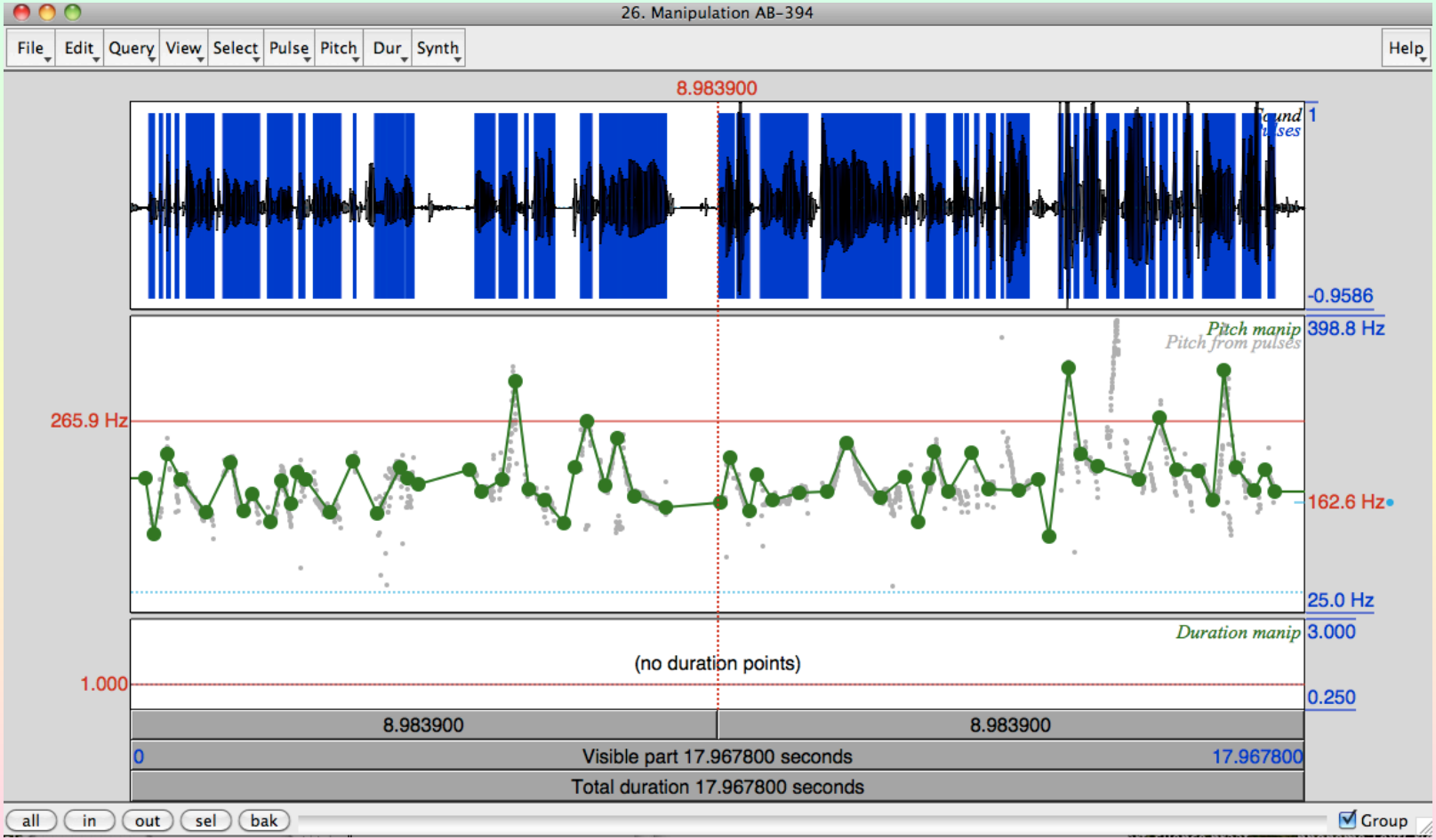

Figure: Extract from recording AB with raw Momel targets.

Momel

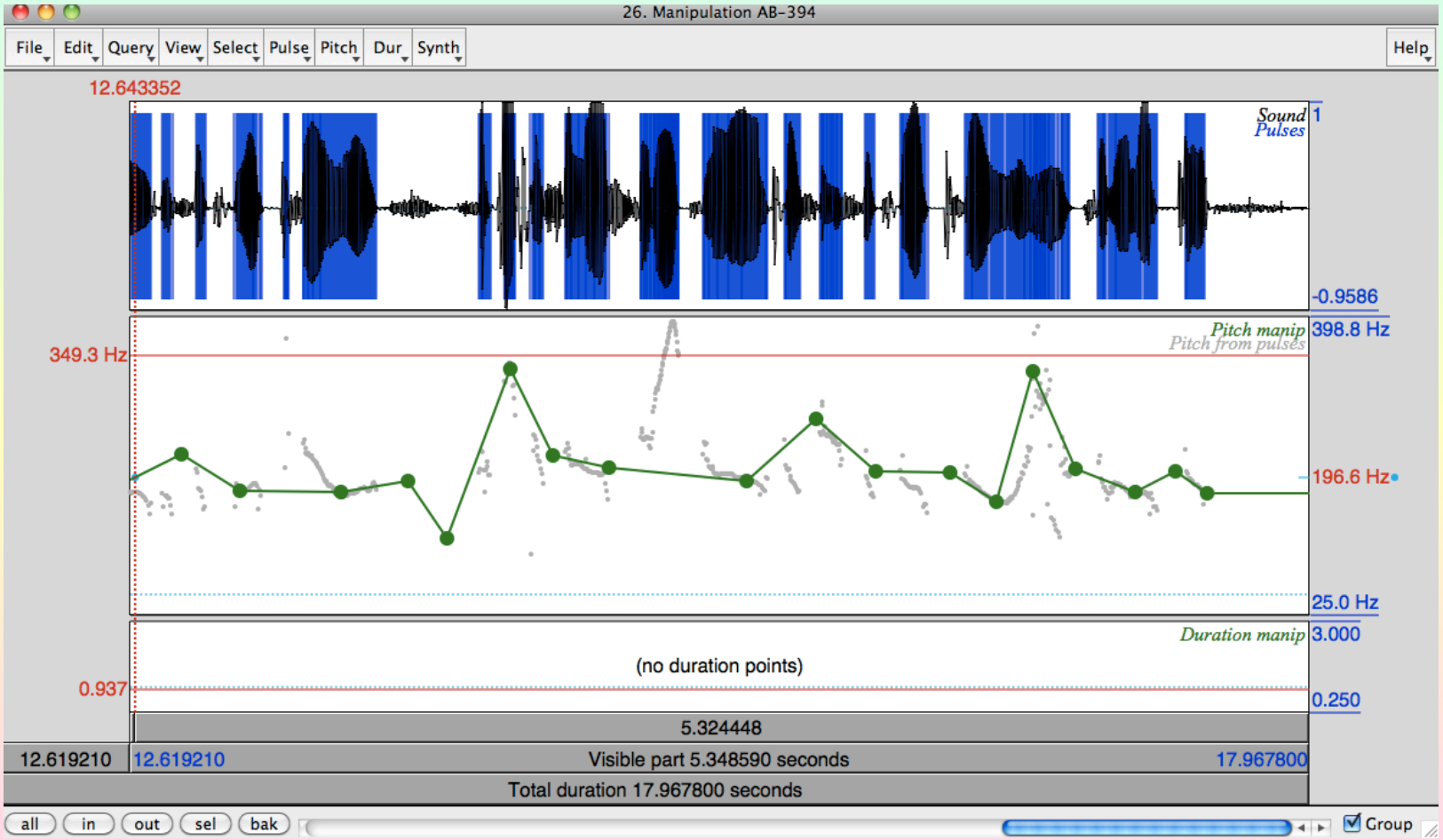

Figure: Detail from recording AB with raw Momel targets.

Momel

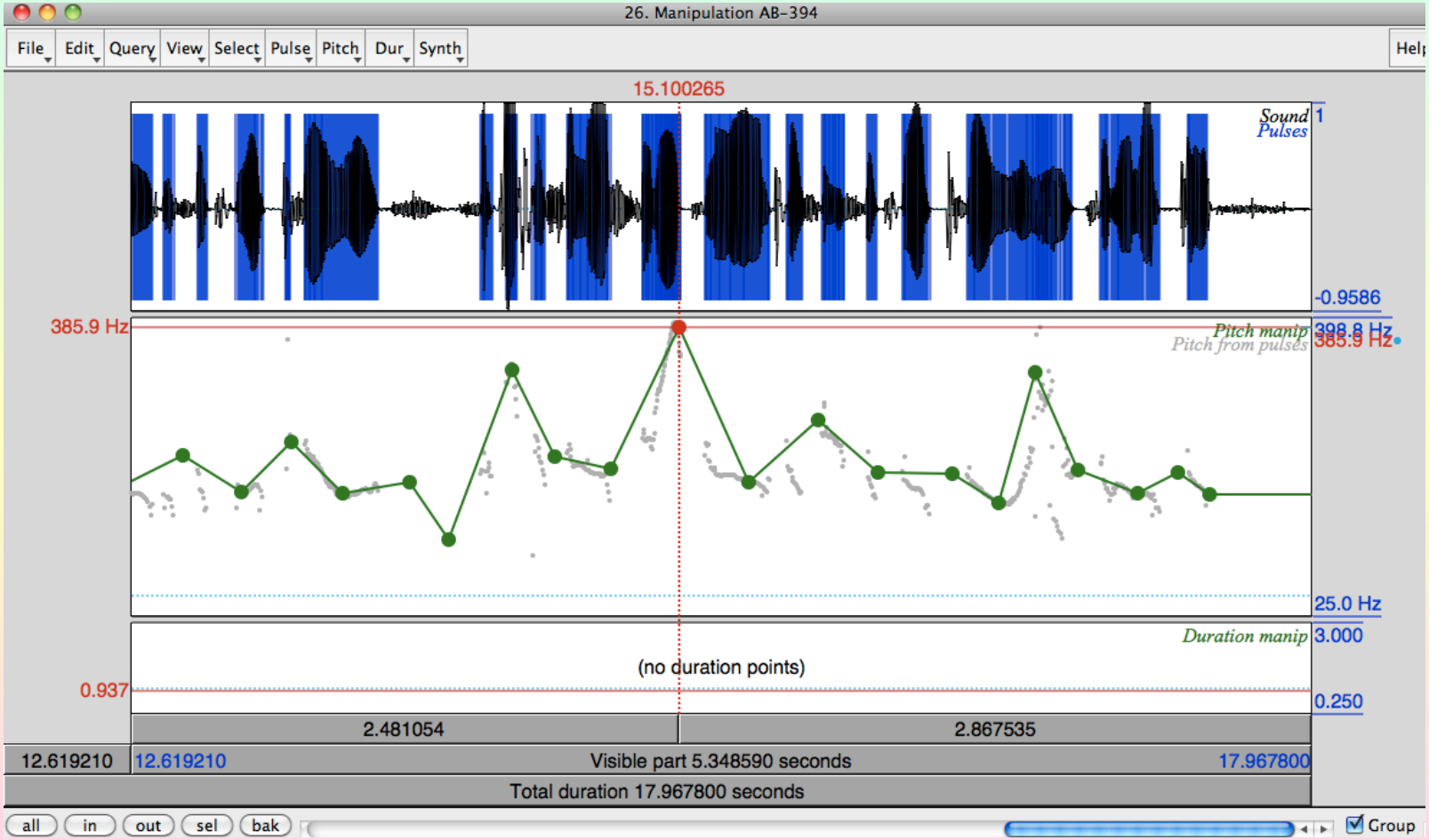

Figure: Detail from recording AB with corrected Momel targets.

Momel

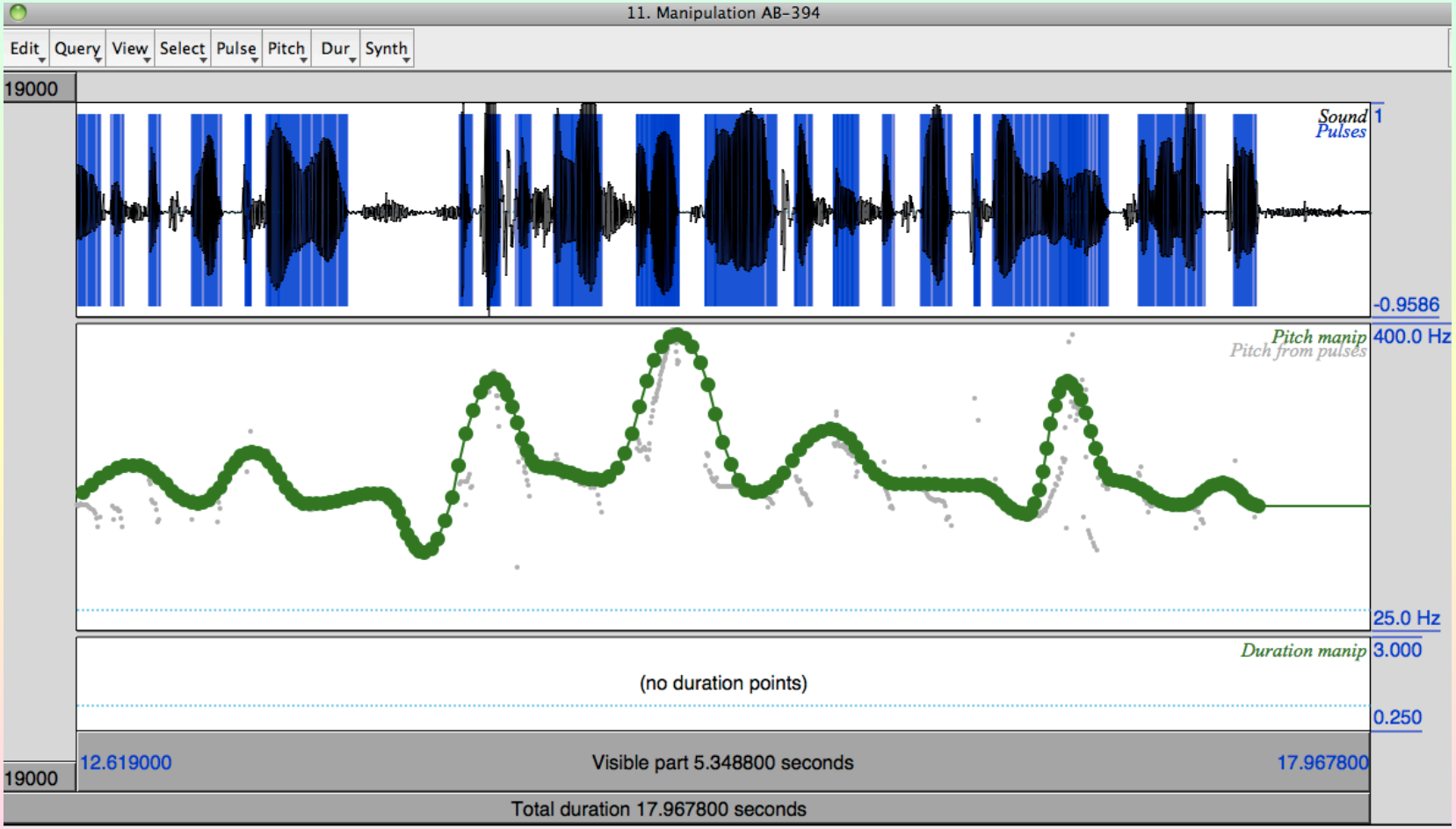

Figure: Detail from recording AB with interpolated Momel curve.

INTSINT

- Target points are defined by two speaker parameters key and $range$

INTSINT

- Target points are defined by two speaker parameters key and $range$
- $T = key * \sqrt{2^{range}}$

INTSINT

- Target points are defined by two speaker parameters key and $range$
- $T = key * \sqrt{2^{range}}$
- $H = \sqrt{P * T}$

INTSINT

- Target points are defined by two speaker parameters key and $range$
- $T = key * \sqrt{2^{range}}$
- $H = \sqrt{P * T}$
- $U = \sqrt{P * \sqrt{P * T}}$

INTSINT

- Target points are defined by two speaker parameters key and $range$
- $T = key * \sqrt{2^{range}}$
- $H = \sqrt{P * T}$
- $U = \sqrt{P * \sqrt{P * T}}$
- etc...

INTSINT

- Target points are defined by two speaker parameters key and $range$
- $T = key * \sqrt{2^{range}}$
- $H = \sqrt{P * T}$
- $U = \sqrt{P * \sqrt{P * T}}$
- etc...
- Optimal coding within target space:

INTSINT

- Target points are defined by two speaker parameters key and $range$
- $T = key * \sqrt{2^{range}}$
- $H = \sqrt{P * T}$
- $U = \sqrt{P * \sqrt{P * T}}$
- etc...
- Optimal coding within target space:

  key mean ± 50 Hz (step: 1)

INTSINT

- Target points are defined by two speaker parameters key and $range$
- $T = key * \sqrt{2^{range}}$
- $H = \sqrt{P * T}$
- $U = \sqrt{P * \sqrt{P * T}}$
- etc...
- Optimal coding within target space:

key	mean ± 50 Hz (step: 1)
span	0.5…2.5 octaves (step: 0.1)

INTSINT

- Target points are defined by two speaker parameters key and $range$
- $T = key * \sqrt{2^{range}}$
- $H = \sqrt{P * T}$
- $U = \sqrt{P * \sqrt{P * T}}$
- etc...
- Optimal coding within target space:

  key mean ± 50 Hz (step: 1)
  span 0.5…2.5 octaves (step: 0.1)

- for each couple $\langle key, span \rangle$ find optimal coding

INTSINT

- Target points are defined by two speaker parameters key and $range$
- $T = key * \sqrt{2^{range}}$
- $H = \sqrt{P * T}$
- $U = \sqrt{P * \sqrt{P * T}}$
- etc...
- Optimal coding within target space:

  key mean ± 50 Hz (step: 1)
  span 0.5…2.5 octaves (step: 0.1)

- for each couple $\langle key, span \rangle$ find optimal coding
- select optimal coding with optimal key and $range$

INTSINT

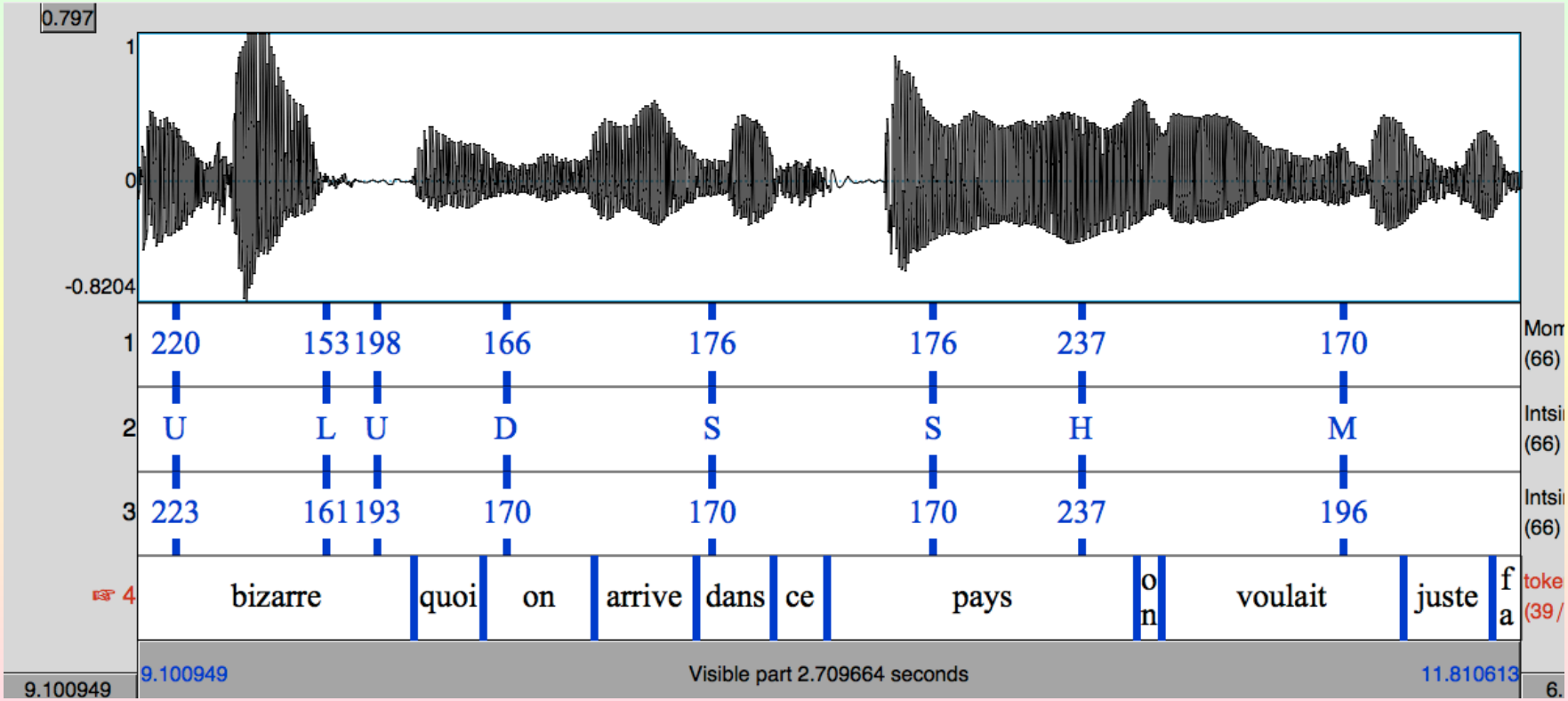

Figure: TextGrid of extract from AB with INTSINT coding.

Perspectives

- Code alignment of INTSINT symbols (with what?)

Perspectives

- Code alignment of INTSINT symbols (with what?)
- Code segmental duration

Perspectives

- Code alignment of INTSINT symbols (with what?)
- Code segmental duration
- With a symbolic representation of prosody, the description of prosody becomes a problem of Natural Language Processing