

Relating ratings of fluency to temporal and lexical aspects of speech

Nivja de Jong & Jan Hulstijn

EALTA, June 2009



AMSTERDAM CENTER
FOR LANGUAGE AND
COMMUNICATION



Overview

- Definition & viewpoints
- Research on aspects of perceived fluency
- This study
 - English data
 - Dutch data
- Discussion

Definition of speaking fluency

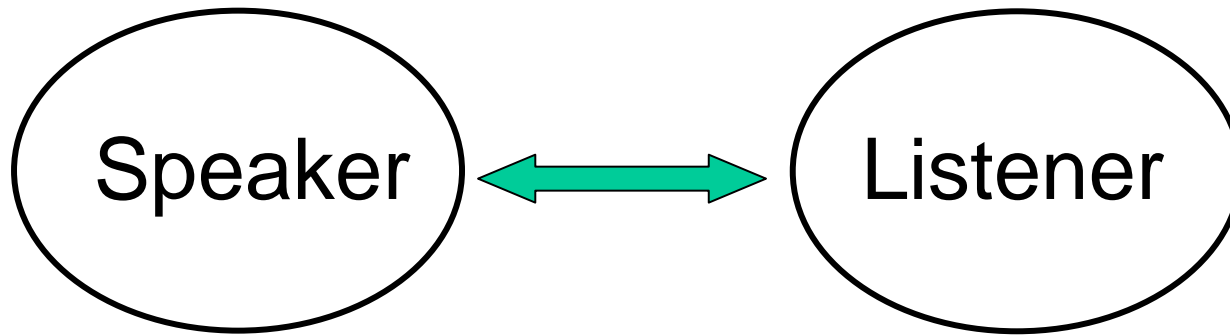
Fluency is automaticity of psycholinguistic processes

Measures of fluency

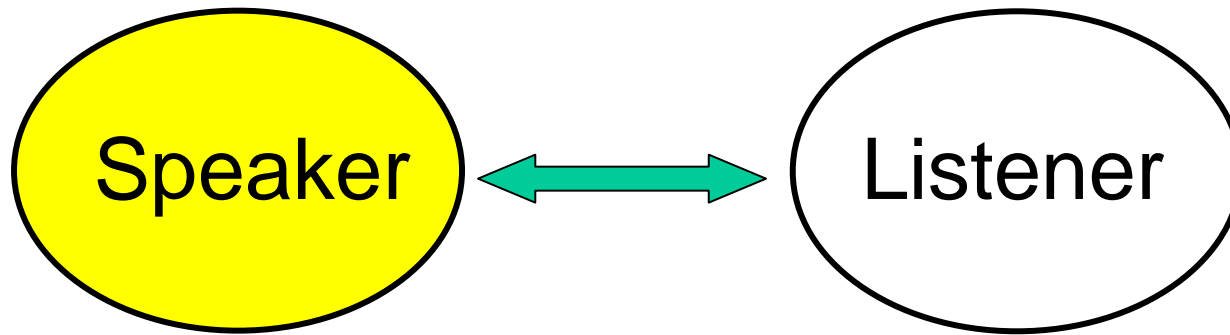
Fluency has a multifaceted nature (Tavakoli & Skehan, 2005):

- Breakdown fluency (e.g., time filled with speech, no. of pauses, filled pauses)
- Speed fluency (e.g., speech rate measured as words per minute, syllables per minute)
- Repair fluency (e.g. false starts, repetitions)

Viewpoints on fluency



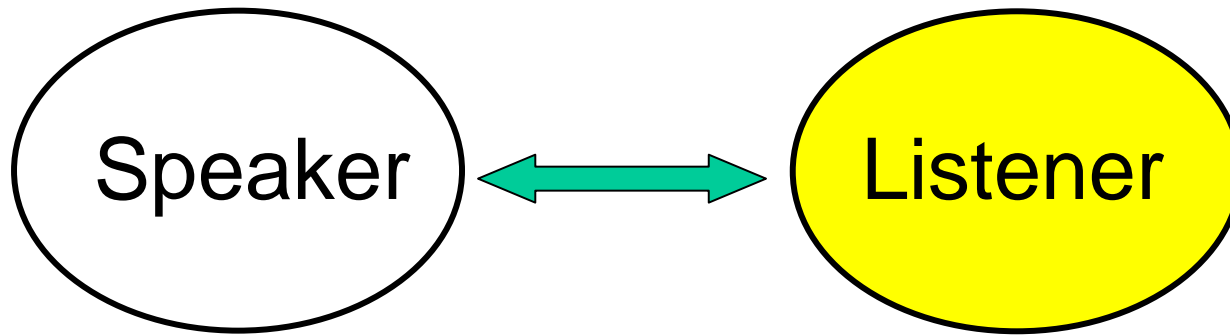
Viewpoints on fluency



What makes speech *be* more or less fluent?

1. Individual characteristics
2. Conceptual planning
3. Formulating

Viewpoints on fluency



What makes speech *sound* more or less fluent?

Fast speech with few pauses, taking into account

1. Individual characteristics
2. Conceptual planning
3. Formulating

Aspects of perceived L1 fluency

Do listeners use information in disfluencies?

- Listeners perceive **pauses at boundaries** to be shorter (Butcher, 1980)
- Low-predictable words preceded by **filled pause**: N400 attenuated (Corley et al., 2007)

Aspects of perceived L2 fluency

Which variables predict raters' fluency?

- **Speech rate**, including pauses (Cucchiarini et al., 2002)
- **Pause / sec** (Derwing et al., 2004)
- **Several fluency measures + accuracy + lexical diversity** (Kormos & Dénes, 2004)

Research Questions

1. Which temporal aspects of speech relate to perceived fluency?
2. Does lexical diversity play a role?
3. Does the pattern differ for lower and higher proficient performances?

Data: English & Dutch speaking performances

English Data

- 1007 performances of Pearson Test of English Academic, various L1 backgrounds
- Task: “detailed descriptions” (monologues)
- 397 performances are transcribed and rated on fluency scale

Dutch Data

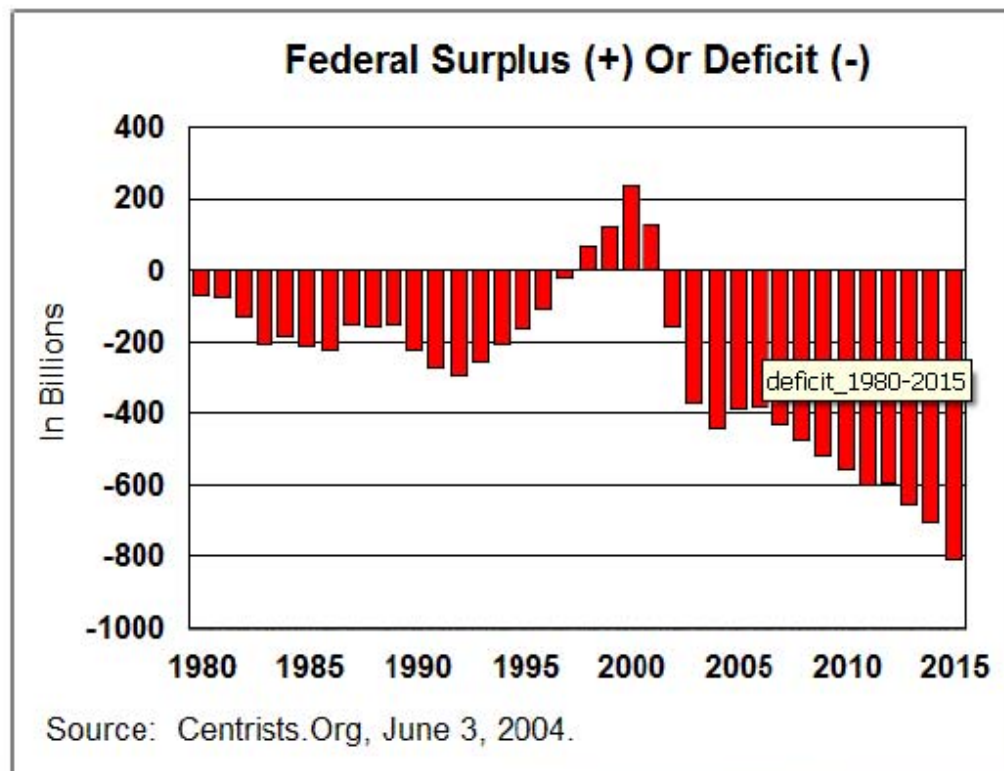
- 1600 performances, various L1 backgrounds
- Tasks: 8 WiSP-tasks (descriptive/persuasive, complex/simple, and formal/informal – all monologues)
- 354 performances rated on CEF spoken fluency scale

English Data

- Example task
- Ratings
- Predictor variables
- Results

Image Description Item

Look at the graph below. In 25 seconds, please speak into the microphone and describe in detail what the graph is showing. You will have 40 seconds to give your response.



Ratings

- 2 raters using oral fluency scale
range 0 – 5
- 2 other raters using adapted CEF Oral Interaction scale
range A2 or below – C2
- Interrater reliability:
 - Fluency-scale: $r = .81$
 - CEF-scale: $r = .78$

Predictor variables

Pausing

- **mean length of pauses** calculated with PRAAT
- **number of pauses** calculated with PRAAT
- **filled pauses percentage** calculated from transcripts

Speech rate

- **syllables per second *excluding pauses*** calculated with PRAAT-script (De Jong & Wempe, 2009)

Lexical diversity

- **Giraud's index: type token ratio** calculated from transcripts

Results: all fluency ratings

N = 397	Correlation	Regression
Mean length of pauses	-.29	-.28
Number of pauses / minute	-.26	-.30
Filled pause percentage	-.18	-.14
Syllables per second	.27	.14
Giraud's index	.34	.26
<i>Explained variance</i>		30%

RQ1:

- Temporal aspects of speech (mainly pausing) are related to perceived fluency

RQ2:

- Lexical diversity also plays a role

RQ3:

- Does the pattern differ for lower and higher proficient performances?

Divide 397 performances into two groups

- 236 CEF scale B1 or below (*mean* Fluency: 2.0, *sd.* 1.1)
- 161 CEF scale B2 or higher (*mean* Fluency: 3.1, *sd.* 1.0)

Results: fluency ratings

Group low proficient (CEF B1 or below)

N = 236	Correlation	Regression
Mean length of pauses	-.32	-.33
Number of pauses / minute	-.26	-.32
Filled pause percentage	-.17	-.16
Syllables per second	.26	.15
Giraud's index	.26	.21
<i>Explained variance</i>		30%

Results: fluency ratings

Group high proficient (CEF B2 – C2)

N = 161	Correlation	Regression
Mean length of pauses	-.10	-.15
Number of pauses / minute	-.16	-.21
Filled pause percentage	-.11	-.12
Syllables per second	.03	-.01
Giraud's index	.18	.16
<i>Explained variance</i>		6%

Conclusion English Data

- Raters' fluency on lower levels of overall proficiency is partly related to global measures of
 - pausing, lexical diversity, and speech rate
- Raters' fluency on higher levels of overall proficiency can only to a very small extent be related to these global measures

Dutch Data

- Ratings
- Predictor variables
- Results

Ratings

- 3 raters using oral fluency scale (CEF)
range 0 – 5
- 4 other raters (from a pool of 12 raters) using scale
for communicative success
range 0 - 30 (>15 successful performance)
- Interrater reliability:
 - Fluency-scale: cronbach's alpha = .76
 - Communicative success: cronbach's alpha = .88 - .90

Predictor variables

Pausing

- mean length of pauses
- number of pauses
- filled pauses percentage

Speech rate

- syllables per second *excluding pauses*

Lexical diversity

- Giraud's index: type token ratio

Results: all fluency ratings

N = 352	Correlation	Regression
Mean length of pauses	-.38	-.28
Number of pauses / minute	-.31	-.26
Filled pause percentage	-.19	.01
Syllables per second	.30	.19
Giraud's index	.55	.40
<i>Explained variance</i>		46%

Results: fluency ratings

Group low proficient (below 15 communicative adequacy)

N = 164	Correlation	Regression
Mean length of pauses	-.46	-.51
Number of pauses / minute	-.11	-.24
Filled pause percentage	.08	.01
Syllables per second	.16	.21
Giraud's index	.18	.01
<i>Explained variance</i>		29%

Results: fluency ratings

Group high proficient (over 15 communicative adequacy)

N = 188	Correlation	Regression
Mean length of pauses	-.10	-.11
Number of pauses / minute	-.26	-.20
Filled pause percentage	-.22	.03
Syllables per second	.26	.21
Giraud's index	.45	.40
<i>Explained variance</i>		27%

Conclusion Dutch Data

- Raters' fluency on lower levels of overall proficiency is partly related to global measures of
 - pausing (mainly pause length), and speech rate
- Raters' fluency on higher levels of overall proficiency is partly related to global measures of
 - number of pauses, speech rate, and lexical diversity

CEF: spoken fluency scale

- C2** Can express him/herself at length with a natural, effortless, unhesitating flow. **Pauses only to reflect on precisely the right words** to express his/her thoughts or to find an appropriate example or explanation.
- B2** Can express him/herself with relative ease. **Despite some problems with formulation resulting in pauses** and 'cul-de-sacs', he/she is able to keep going effectively without help.
- A2** Can construct phrases **on familiar topics** with sufficient ease to handle short exchanges, despite very noticeable **hesitation and false starts**.

Discussion

- What is L2 fluency?
- What is L2 fluency in context of L2 speaking tests?
 - Raters, fluency scale?
 - Objective measures?
 - Comparison to L1 fluency?
 - L1 fluency research
 - Individual characteristics?

Thanks

- English data:

Pearson plc

- Help with PRAAT-scripts:

Ton Wempe

- Members of WiSP-team

Rob Schoonen

Margarita Steinel

Arjen Florijn

Questions?

and future research

Include more global measures taken from the transcripts: item difficulty, lexical profile, grammatical profile, accuracy

Include local analyses of fluency

CEF: spoken fluency scale

B1 Can keep going comprehensibly, **even though pausing for grammatical and lexical planning** and repair is very evident, especially in longer stretches of free production.

Can make him/herself understood in short contributions, even though **pauses, false starts and reformulation** are very evident.

A2 Can construct phrases **on familiar topics** with sufficient ease to handle short exchanges, despite very noticeable **hesitation and false starts**.

A1 Can manage very short, isolated, mainly pre-packaged utterances, with **much pausing to search for expressions**, to articulate less familiar words, and to repair

Possible explanations

- Variances of Fluency ratings differ between High and Low proficient group
- Variances of predictors differ between High and Low proficient group
- Fewer performances (161 vs 236) in High group

Possible explanations

- Variances of Fluency ratings differ between High and Low proficient group
- Variances of predictors differ between High and Low proficient group
- Fewer performances (161 vs 236) in High group
- Fluency at higher levels of proficiency is related to other speech characteristics

Definitions of speaking fluency

Fluency is the combination of speed and smoothness or effortless

Fluency is smoothness in terms of temporal, phonetic and acoustic features

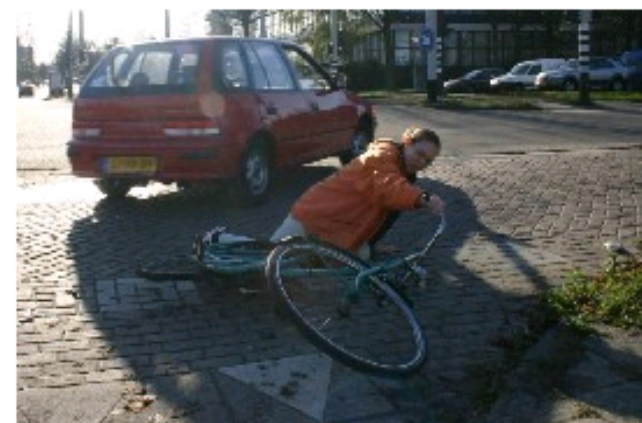
Fluency is automaticity of psychological processes

Koponen & Rigggenbach (2000): it is not possible to isolate a single unitary concept of fluency

Hieronder ziet u wat u een maand geleden op straat hebt gezien.

De rechter vraagt u om precies te beschrijven wat u heeft gezien.

- Begin uw woorden met "Geachte rechter..."
- Vertel dat u het ongeluk heeft gezien
- Beschrijf in detail wat er gebeurd is

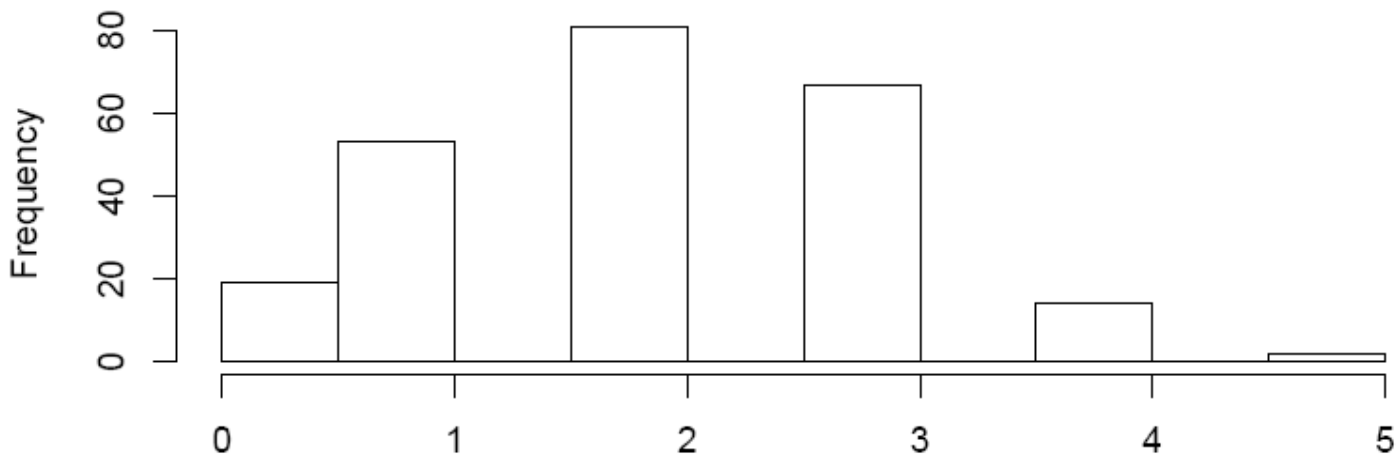


Bereid nu voor wat u wilt gaan zeggen

Conclusion: L2 perceived fluency

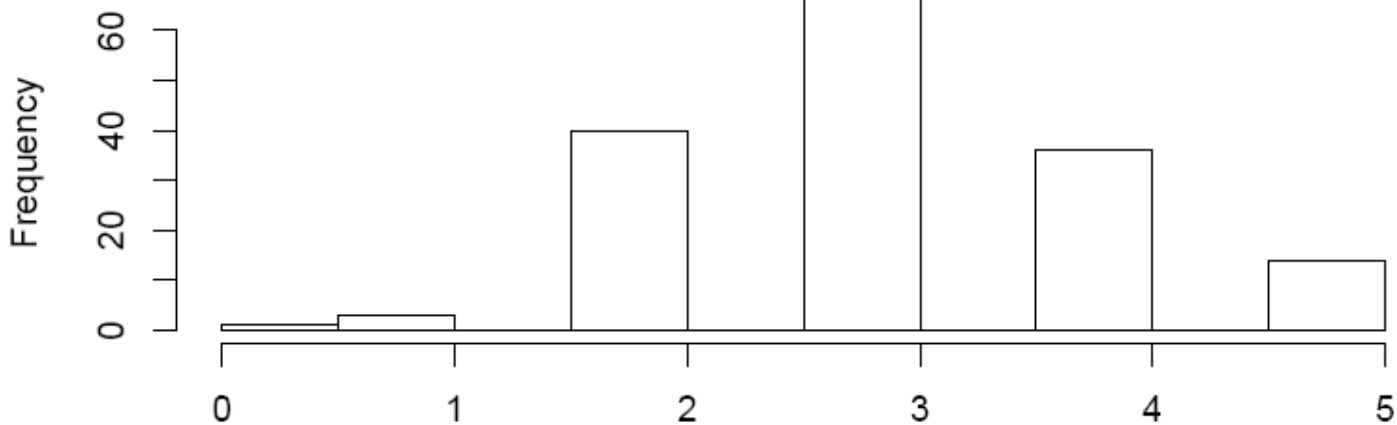
- Raters' fluency is related to more than just temporal measures
- Raters (partly) do as they are told
- Raters' fluency is probably not the best starting point

Histogram of FLU (CEF A2 - B1)



N = 236

Histogram of FLU (CEF B2 - C2)



N = 161

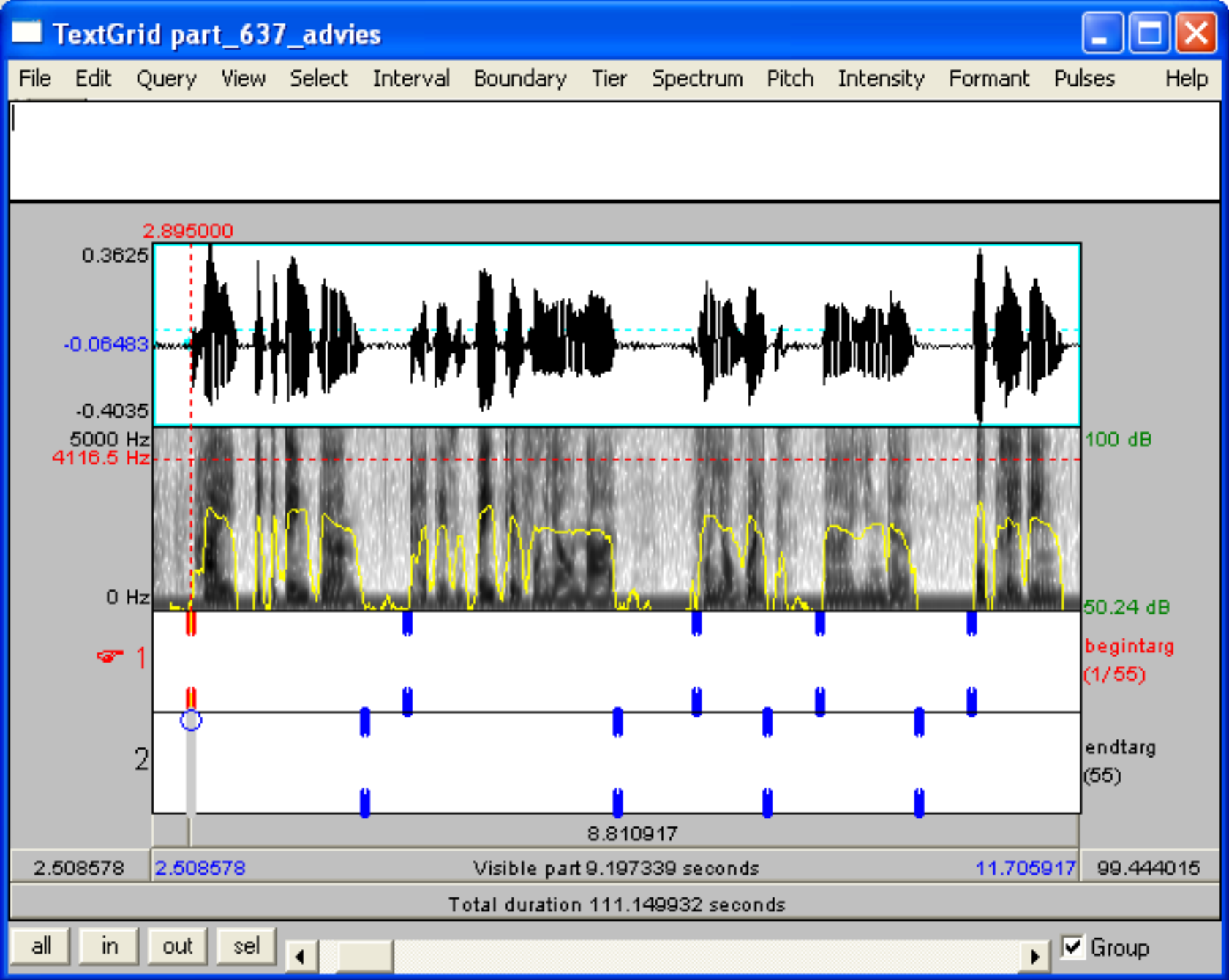
Automatic measures of fluency

Scripts written in PRAAT

- Measuring speech/silence: if sound is voiced, determine beginning of speech with intensity (dB) above a certain threshold

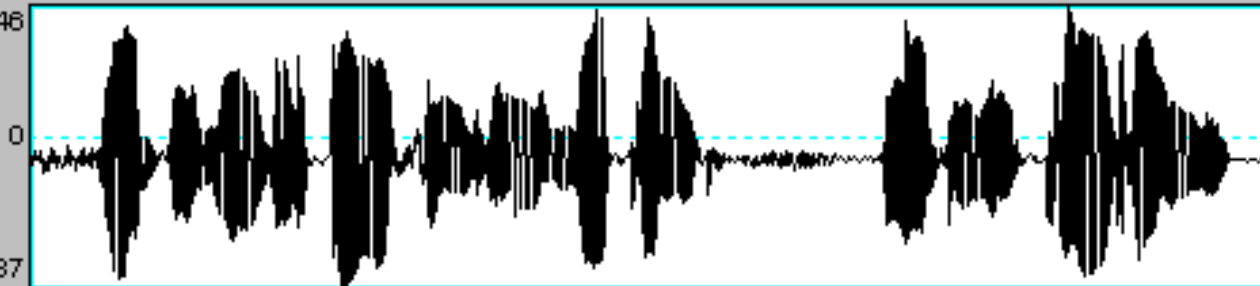
De Jong & Wempe (in press *Behavior Research Methods*)

- Measuring syllable nuclei: if intensity (dB) is above a certain threshold, and extent of previous dip in intensity is above certain threshold, and if sound is voiced



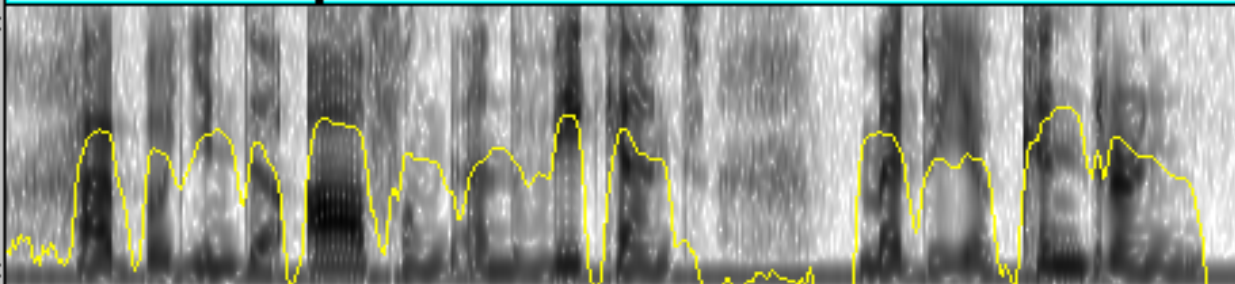
2.209

0.3846



-0.4437

5000 Hz



100 dB

0 Hz

50 dB

1

60 61 62 63 64 65 66 67 68 69 70 71 72 73

syllables (228)

33.212600

Visible part 4.417845 seconds

37.630444

33.212600

73.519487

Total duration 111.149932 seconds

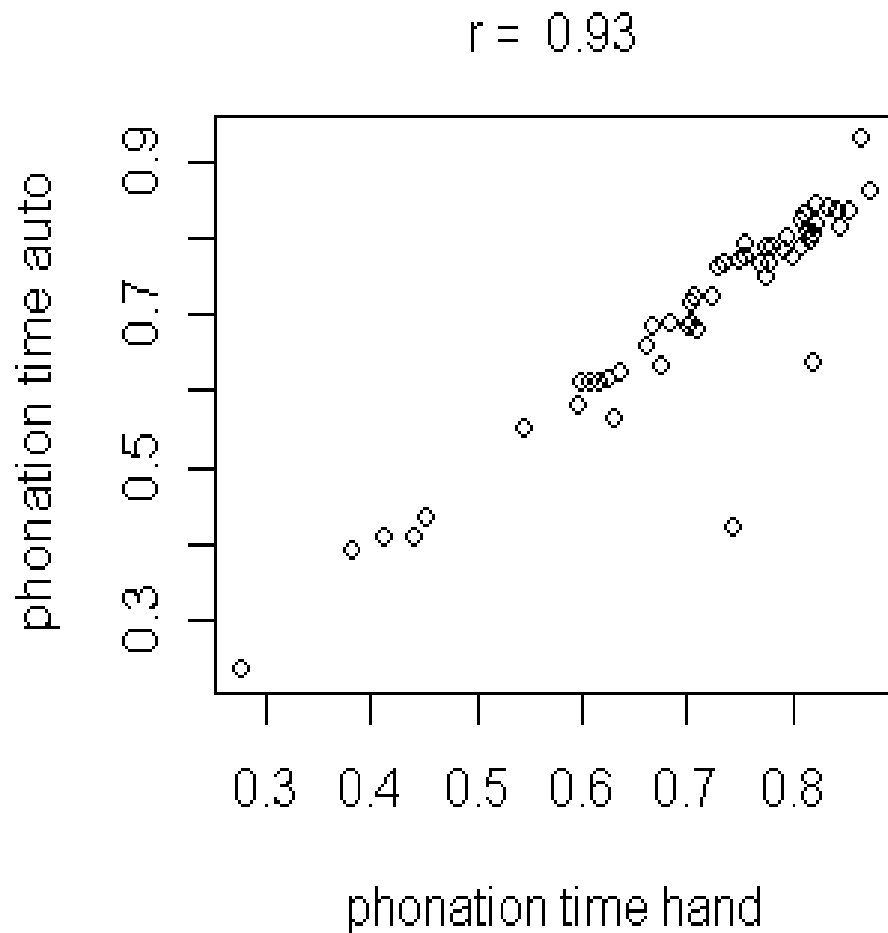
all in out sel



Group

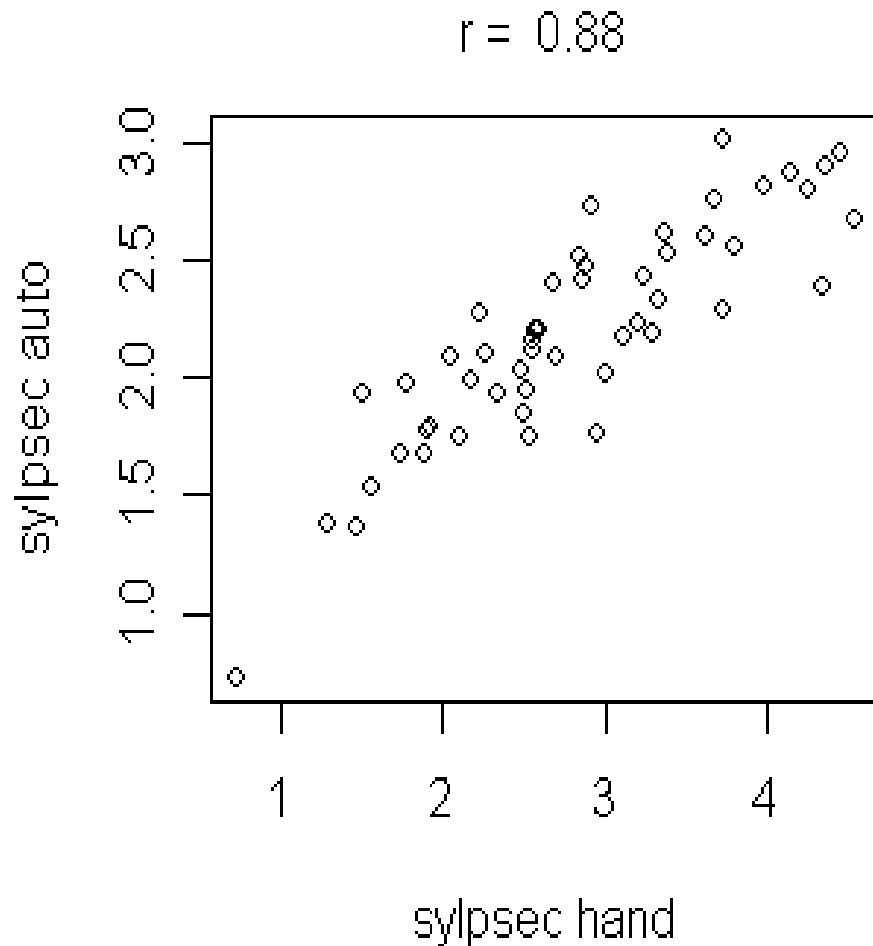


Validating automatic measurement of phonation time ratio



Correlation
between hand
and automatic
measure: $r = 0.93$

Validating automatic measurement of syllables per second



Correlation
between hand and
automatic
measure: $r = 0.88$

Results: phonation time ratio

	Correlation	Regression
Vocabulary knowledge	.14	-.04
Grammar knowledge	.15	.11
Lexical retrieval speed	-.17	-.08
Articulation latency	-.07	-.03
Pronunciation duration	.26	.25
Sentence building speed	-.14	-.08
Extraversion	.14	.11

Results: filled pauses percentage

	Correlation	Regression
Vocabulary knowledge	-.36	-.27
Grammar knowledge	-.22	.19
Lexical retrieval speed	.33	.16
Articulation latency	.12	-.01
Pronunciation duration	.00	.04
Sentence building speed	.40	.21
Extraversion	-.20	-.17

Results: syllables per second

	Correlation	Regression
Vocabulary knowledge	.50	.41
Grammar knowledge	.40	-.05
Lexical retrieval speed	-.28	.03
Articulation latency	-.16	-.06
Pronunciation duration	-.06	-.07
Sentence building speed	-.47	-.22
Extraversion	.06	.04

Fluency Scale © Ordinate Corporation

- 5 **NATIVE-LIKE Fluency.** Candidate utterance exhibits smooth native-like rhythm and phrasing, with no hesitations, repetitions, false starts, or non-native phonological simplifications.
- 4 **ADVANCED Fluency.** Candidate utterance has acceptable rhythm, with appropriate phrasing and word emphasis. Utterances have no more than one hesitation, repetition or false start. There are no significantly non-native phonological hesitations.
- 3 **GOOD Fluency** Candidate speech has acceptable speed, but may be somewhat uneven. Long utterances may exhibit more than one hesitation; but most words are spoken in continuous phrases. There are few repetitions or false starts per utterance. Speech has no long pauses, and does not sound staccato.

Fluency Scale © Ordinate Corporation

- 2 INTERMEDIATE Fluency.** Candidate speech may be uneven or somewhat staccato. Utterance (if ≥ 6 words) has at least one smooth 3-word run, and no more than two or three hesitations, repetitions or false starts. Speech may have one long pause, but not two or more.
- 1 LIMITED Fluency.** Candidate speech has irregular phrasing or sentence rhythm. Poor phrasing, staccato or syllabic timing, and/or multiple hesitations, repetitions or false starts render the spoken performance notably uneven or discontinuous. Long utterances may have one or two long pauses and may have inappropriate sentence-level word emphasis.
- 0 DISFLUENT** Candidate speech is slow and seems labored, with little discernable phrase grouping and with multiple hesitations, pauses, false starts and/or major phonological simplifications. In an utterance, most words are isolated and there may be more than one long pause.

Overall Test Design

Item-type	Main Skills	Activity	# Items
Single-answer, multiple choice	Reading & Vocabulary	Select the correct answer	3
Multi-answers, multiple choice	Reading & Vocabulary	Select the correct answers	3
Highlighting	Reading & Referential	Highlight the text	3
Drag and drop	Reading & Referential	Jigsaw text: drag and drop	3
Drag and drop, gaps	Reading, Vocabulary & Referential	Gap fill: drag and drop	3
Summary of written text	Reading & Writing	Write a one-sentence summary	3
Summary of spoken text	Listening & Writing	Write a one-paragraph summary	3
Essay	Critical thinking & Essay writing	Write a 200-300 word essay	1
Dictation	Listening for detail, Writing	Transcribing	3
Listening, fill in the gaps	Listening, Writing & Vocabulary	Close listening	3
Writing, fill in the gaps	Reading, Writing & Vocabulary	Close reading, multiple choice	3
Read out loud	Reading & Speaking	Reading a text aloud	3
Repeating	Listening & Speaking	Listening and repeating sentences	20
Detailed descriptions	Critical thinking & Speaking	Describing graphs and images	2
Retelling	Listening & Speaking	Retelling a lecture	2
Short Questions	Listening, Speaking & Vocabulary	Replying to questions	20
Single-answer, multiple choice	Listening & Vocabulary	Select the correct answer	3
Multi-answers, multiple choice	Listening & Vocabulary	Select the correct answers	3
Listening word gap	Listening & Vocabulary	Passage completion: select the word	3
Highlighting errors	Listening for detail & Reading	Listen and read: highlighting errors	3
Highlighting answers	Listening, Inferencing & Reading	Listen and read: select the summary	3