Interaction Between Fluency Measures and Task Characteristics

April Ginther, Purdue University
Slobodanka Dimova, Copenhagen University
Soohwan Park, Purdue University

Introduction

This study extends previously published findings (Ginther, Dimova, and Yang, 2010) examining the relationships among temporal measures of fluency and holistic scores on a semi-direct measure of oral English proficiency, the Oral English proficiency Test (OEPT).

While studies examining fluency have established that speech rate (SR), mean syllable per run (MSR), and number of silent pauses (NSP) are reasonable and (relatively) easily measured proxies for fluency, each of these variables represents different underlying aspects of the phenomenon, and the selection of one over the other has different theoretical and practical implications. For example, the finding that SR is strongly and positively related to holistic scores of oral proficiency might lead an examinee to conclude that receiving a higher score on an oral proficiency test is simply a matter of speaking faster, but speaking faster may be easier said than done. While this conclusion seems reasonable, examinees at higher levels of proficiency also pause less often and, at the same time, they produce longer runs between silent pauses. It may be the case that SR is actually a function of MSR; that is, faster measured speech rates are a function of an examinee's ability to produce longer runs with fewer silent pauses. In other words, MSR may enable SR, so learners and teachers might be best served by focusing instruction on students' development of the production of multi-syllabic runs with limited and appropriately placed silent pauses, rather than speech rate.

Secondly, It would be interesting to know whether the ability to produce longer runs is influenced by task characteristics. If a task presents more language and information that can be used by a speaker in responding to a prompt, it may be the case that fluency will be facilitated, and reflected in longer runs, if and when examinees make use of the language and information presented in a prompt.

Finally, as Ginther, Dimova, and Yang (2010) demonstrated, SR and MSR are strongly and positively correlated (.85); however, with respect to further investigation, SR presents a kind of investigative dead end. There's not a lot that can be done with SR once proficiency level descriptives have been established. MSR, on the other hand, provides a rich artifact when speech samples are transcribed, easily allowing further investigation. A logical next step is to examine whether an examinee's ability to produce runs bounded by 'expected' pauses, i.e., runs that can be described as syntactically 'well-formed' differ across proficiency level. For these reasons, we have selected MSR as the dependent variable of interest.

Oral English Proficiency Test (OEPT)

- A semi-direct, computer-administered exam of oral English proficiency
- Administered to 500 prospective international teaching assistants (ITAs) at Purdue University every academic year
- 12 items, requiring performance of tasks associated with a university teaching context
- A 4-point holistic scale (levels 3-6), 2 raters
  - Normally distributed w/ a pass rate of 50%
  - Alpha .99; SEM .11; single factor
  - Inter-rater reliability: .78 - .84

Item	Title	Task	Presentation
1	Area of Study	Describe interests	Text/audio
2	Newspaper Headline	Explain an opinion	Text/audio
3	Compare and Contrast	Compare information and explain a decision	Text
4	Pros and Cons	Consider consequences and explain a decision	Text/Audio
5	Give Advice	Consider a problem and offer a solution	Text/Audio
6	Bar Chart	Describe and interpret graph	Graph
7	Line Graph	Describe and interpret graph	Graph
8	Voicemail Message	Relay information	Audio only
9	Conversation	Summarize conversation	Audio only
10	Short Lecture	Summarize short lecture	Audio only
11	Read Aloud 1	Read English text intelligibly	Text
12	Read Aloud 2	Read English text intelligibly	Text

Table1: OEPT Structure

Focal Task Characteristics

Task characteristics have been examined in terms of task dependence of response processing and production. In other words, we selected the focal tasks for this study by analyzing the degree to which the elicited response repeated the information presented in the task prompt.

Task1: Read aloud (RAL). It does not require production of new language but simple oral reproduction of the written text presented in the prompt.

Task 2: Compare and contrast (CNC). It requires some repetition of the text presented in the prompt (e.g. a list of characteristics for each item to compare), but it also requires independent speech production of new ideas.

Task 3: News passage (NP). The prompt contains only one sentence, so examinees have to construct a response that doesn't rely on the text presented in prompt.

Research Questions

1.When item responses to the News passage item are compared to item responses to the Compare and Contrast and the Read Aloud items, would fluency, as measured by mean syllables per run, be facilitated when (a) more information is provided in the prompt (Compare and Contrast) or (b) when the response is entirely prompt dependent (Read Aloud)?

2.Across all three items, in addition to mean syllables per run and the number of pauses, when pause placement is taken into account, would there be a tendency for examinees at lower levels of proficiency to pause more often at unexpected locations and for higher proficiency examinees to pause more often at expected locations?

Participants

- A total of 450 speech samples
- Chinese, Hindi, and English speakers
- Chinese and Hindi (Age=19-34; Male=70% and Female=30%)
- English (Age=20-71, Male=28% and Female=72%)
- College of Science, College of Engineering, and Liberal Arts

	Chinese 3	Chinese 4	Chinese 5	Hindi 5	Hindi 6	English 7	Total
NP	25	25	25	25	25	25	150
CNC	25	25	25	25	25	25	150
RAL	25	25	25	25	25	25	150

Table 2: Number of participants by L1, level, and item type

Variables Examined in the Study

Mean Syllables per Run (MSR) = number of syllables divided by number of runs in a given speech sample

Runs = number of syllables produced between two silent pauses

Silent pauses = pauses equal to or longer than 0.25 seconds

Expected pauses = pauses occurring

- After period (or between sentences)
- After subject clause (or before verb or auxiliary verb)
- Before preposition
- Before conjunctions (and, or, and but)
- Before complementizer (if and that)
- Before relative pronoun and adverb
- At punctuations (comma, colon, and semi-colon)
- Before 'and' and 'because'

Unexpected pauses = pauses placed other than expected pausing positions

Analysis

PRAAT was used to extract and transcribe temporal variables

1. the length of each pause in seconds
2. the number of syllables uttered between pauses;
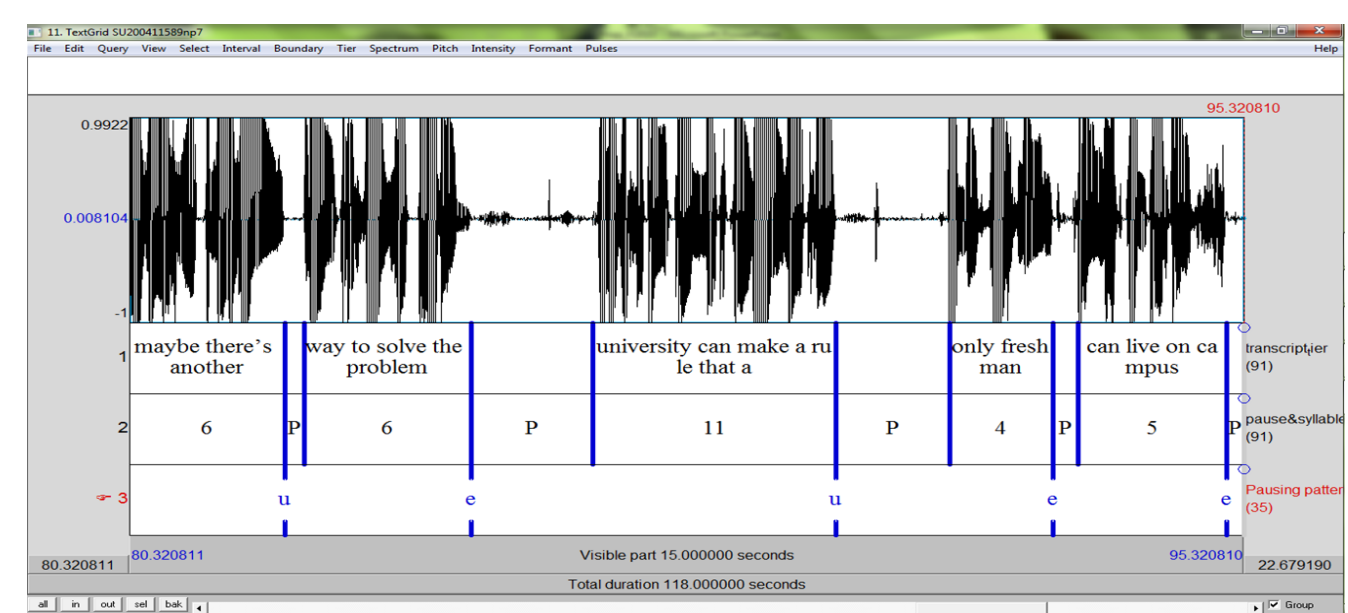3. the length of speech time in seconds between the pauses.

Figure 1: Text grid in PRAAT
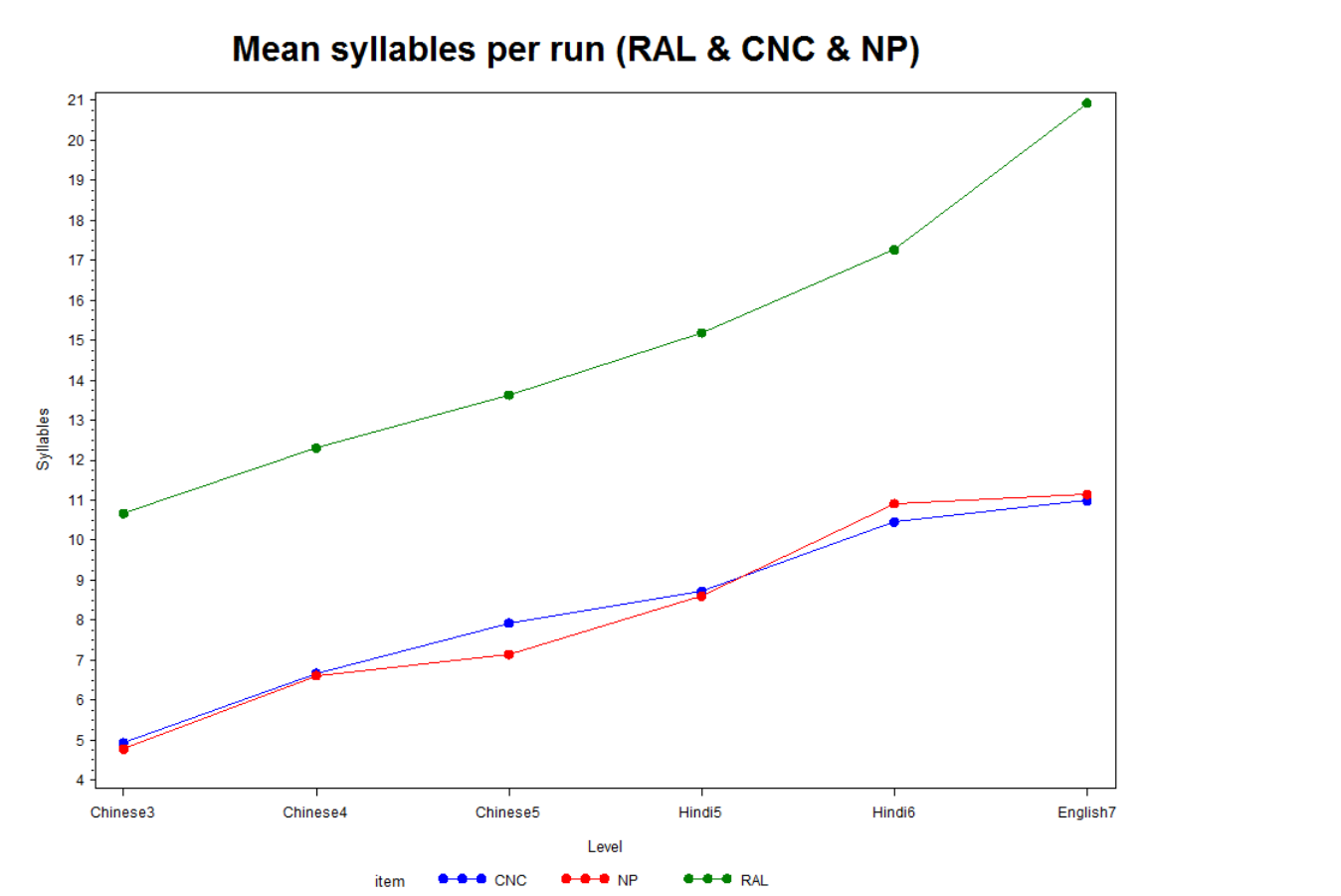
Results

Research Question 1

- A proficiency level by prompt (6x3) ANOVA produced a significant interaction when mean syllables per run was selected as the DV

- Fluency facilitated by prompt characteristics only on the RAL item where performance was completely prompt dependent

Item	Level	N	Mean	SD	Min	Max	Lower 95% CI	Upper 95% CI
NP (.75)	Chinese 3	25	4.78	1.18	3.28	7.54	4.30	5.27
	Chinese 4	25	6.62	1.52	4.03	9.83	5.99	7.25
	Chinese 5	25	7.15	1.50	5.26	10.81	6.53	7.76
	Hindi 5	25	8.60	2.74	5.04	15.75	7.47	9.73
	Hindi 6	25	10.92	3.01	6.12	19.17	9.67	12.16
	English 7	25	11.14	3.11	5.21	17.53	9.86	12.42
CNC (.71)	Chinese 3	25	4.94	1.39	3.33	9.31	4.37	5.51
	Chinese 4	25	6.68	1.99	4.09	11.79	5.86	7.50
	Chinese 5	25	7.93	2.58	5.07	18.82	6.87	9.00
	Hindi 5	25	8.72	2.71	5.35	17.14	7.60	9.84
	Hindi 6	25	10.46	2.23	6.51	15.06	9.54	11.38
	English 7	25	10.99	3.17	5.13	17.88	9.68	12.30
RAL (.65)	Chinese 3	25	10.67	4.13	4.50	20.11	8.97	12.38
	Chinese 4	25	12.31	2.37	8.50	16.68	11.33	13.29
	Chinese 5	25	13.64	3.31	9.40	22.25	12.27	15.00
	Hindi 5	25	15.19	4.21	6.48	25.57	13.45	16.93
	Hindi 6	25	17.26	5.05	10.63	34.00	15.18	19.34
	English 7	25	20.93	4.67	13.83	33.30	19.01	22.86

Table 3: Descriptive statistics for mean syllables per run

Source	Df	SS	MS	F
Item	2	4567.11	2283.56	247.98**
Level	5	2930.21	586.04	63.64**
Level × Item	10	215.55	21.55	2.34*
Errors	432	3978.09	9.21	
Total	449	11690.96		

Table 4: Model for mean syllables per run (RAL, CNC, NP)

Mean syllables per run (RAL & CNC & NP)

Results

Research Question 2

- As proficiency levels increase, runs associated with all prompts were not only longer but also increasingly 'well-formed' *vis-à-vis* pause placement.

*Mean syllables per run with expected and unexpected pauses: Chinese 3 example*

I think
E university students
E (should be involved in programs that involve interaction with elementary schools children)
E because I think
E university students
E have different social
U social background
E with
E elementary schools children
E the society is chan—
U is change
U the society
E keep changing
U keep changing and…

Mean syllables per run with expected and unexpected pauses: Chinese 5 example

E I think in every state of life there are always problems to face
E so
E like the
U elementary stu—
U school students they may have their
U their problems of life and
E sometimes they really need some instruction to face
E these problems
E and
E for the college students
E they will also have their problems such as the
U their study

Mean syllables per run with expected and unexpected pauses: Hindi 6 example

E I do not believe I believe there are a number of issues or queries or questions that children have that they're apprehensive to talk about to their chil– with their parents
E a number of these
E or these issues or questions relate
E to smoking to sex to drinking
E and these are issues that children don't necessarily

Number of expected pauses per minute (RAL & CNC & NP)

Number of unexpected pauses per minute (RAL & CNC & NP)
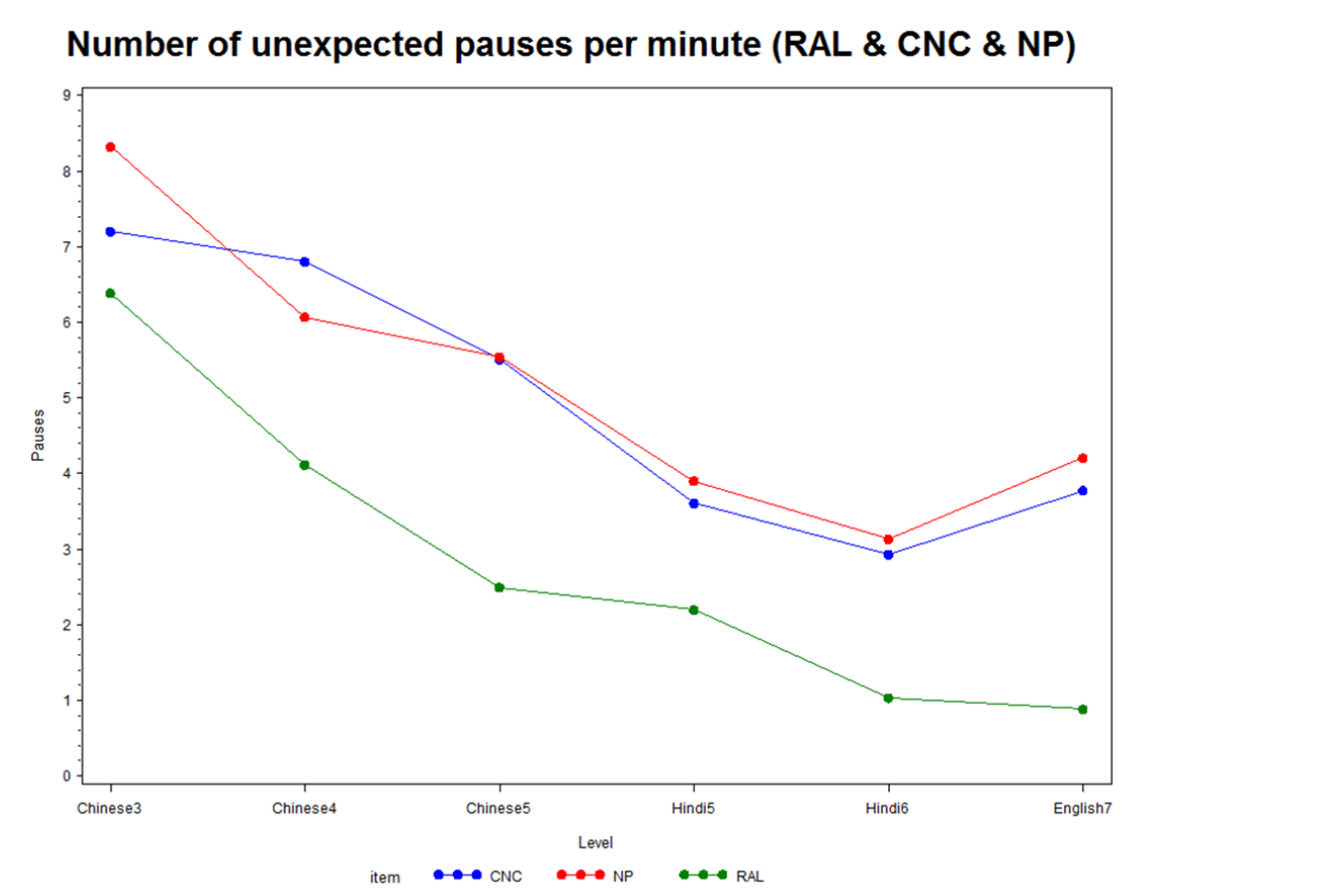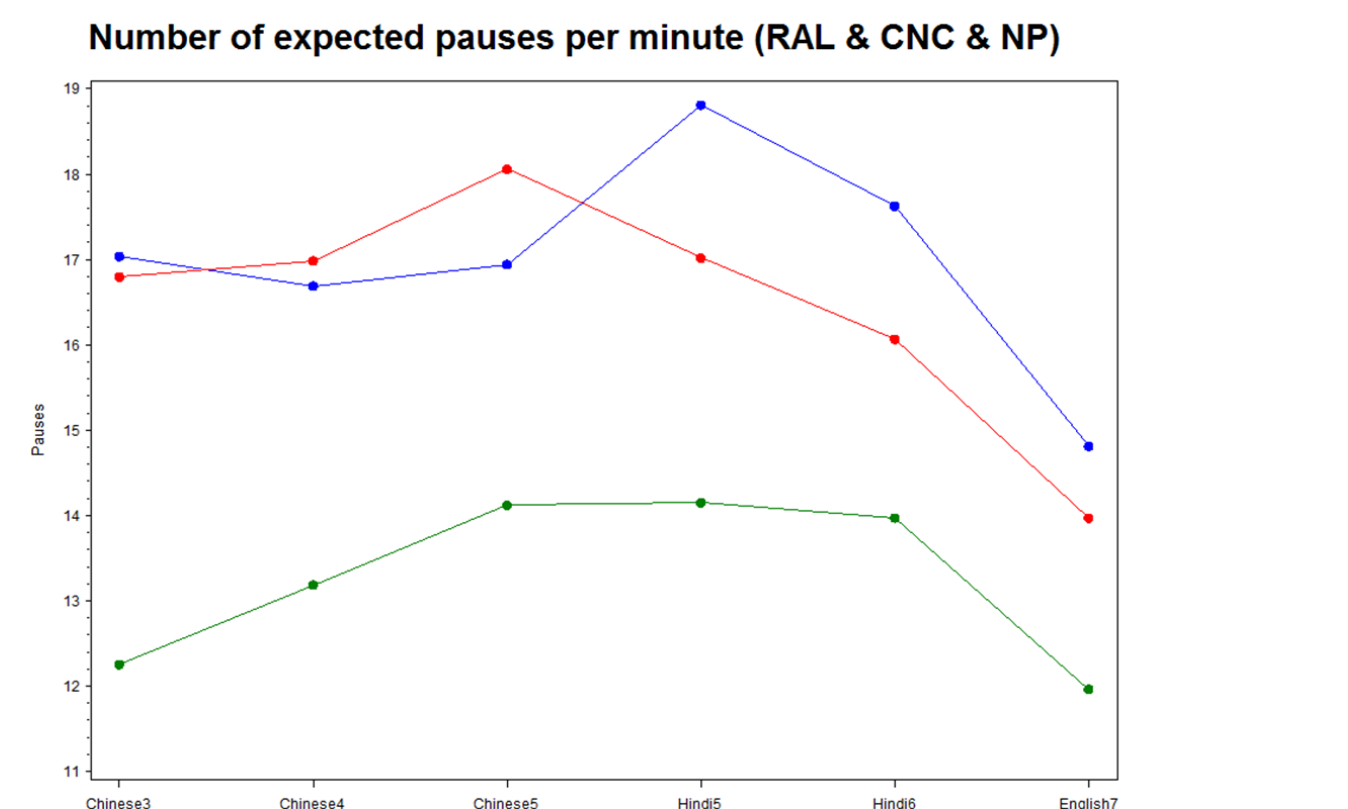
Conclusions

- Contrary to expectations, task characteristics have no effect on fluency unless the task is extremly prompt dependent as in RAL.

- The unexpectedness with regard to the expected/unexpected pauses may not at all be a function of the well-formedness of the run but rather the interruptions caused by restarting and/or abandoning a particular path.

Reference
Ginther, A., Dimova, S., and Yang, R. (2010). Temporal measures of fluency and holistic ratings of spoken English with implications for automated scoring. *Language Testing*, 27, 379-399.