The role of fluency and rhythm in the assessment of task delivery

Takayuki Konishi (Kobe University), Yoichi Mukai (Vancouver Island University), Shungo Suzuki (Nagoya University), Kyoko Hitomi (Waseda University), Ryuki Matsuura (Carnegie Mellon University) & Katsunori Kanzawa (Kyoto Institute of Technology)

Email Takayuki | Email Yoichi

Second Language Research Forum 2025
September 26, Northern Arizona University

1. Introduction

Background

- Oral fluency is considered essential in the learning, teaching, and assessment of speaking skills. [1]
- Fluency is a complex and dynamic construct, leading to differing definitions. [2]
- In L2 research, fluency is narrowly defined as temporal features of speech, whereas in assessment it is often conflated with rhythm and expanded to include delivery.

Purpose

- This study examines the relative importance of temporal and rhythmic features in the assessment of task delivery.

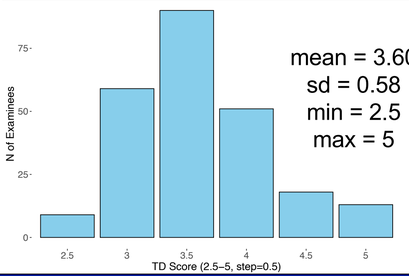
2. Data – KIT Speaking Test Corpus

The KIT Speaking Test	Our study
9 tasks: 3 photo description (Q1-3), 2 conversation summary (4, 6), 2 opinion (5, 7), and 2 structured speech (8, 9), 45-60 sec each	1 semi-structured task (Q4) and 3 spontaneous tasks (3, 5, 9).
Task Achievement (TA) and Task Delivery (TD) **scores** (by 1 native & 1 non-native rater)	**TD scores** only (which primarily evaluates fluency)
574 **participants** with scored speech samples	Top 60 only (to avoid excessive disfluency)

[5] Mostly confident AND time used well with no intrusive pauses, hesitations, or repetitions

[1] Gives up OR meaning not conveyed with delivery problems (slow speech, pauses, hesitations, repetitions)

[0] Does not start the task



mean = 3.60
sd = 0.58
min = 2.5
max = 5

Scan for further details

Link to the corpus

Segmentation

- 4 tokens x 4 tasks x 60 participants (to exclude fillers and disfluency to calculate rhythm scores) = 960 tokens, mean duration = 2.31 ms, sd = 0.93.
- Annotation by Montreal Forced Alignment [3] + manual modification.

3. Analysis

Variables of interest

speed	**articulation rate** (# of syllables / speech dur)
repair	**disfluency ratio** (freq of disfluency / syllable)
pause	mid-clause pause ratio/duration, end-clause pause ratio/duration, filled pause ratio
rhythm	ΔV/C, %V, VarcoV/C, nPVI-V/C, rPVI-C

- Calculated with automatic annotation systems [4].
- To avoid multicollinearity, the number of factors were reduced by principal component analyses.

Pause-related factors (Fig. 1)

	duration	frequency
mid-caluse pause ratio	0.342	-0.645
end-clause pause ratio	0.447	-0.070
mid-caluse pause dur	0.574	-0.123
end-clause pause dur	0.557	0.325
filled pause ratio	-0.209	-0.677

Rhythm-related factors (Fig. 2)

	Var-C	Var-V		Var-C	Var-V
ΔC	-0.518	0.006	ΔV	-0.049	0.573
VarcoC	-0.449	0.045	%V	0.215	0.388
nPVI-C	-0.458	-0.021	VarcoV	-0.031	0.532
rPVI-C	-0.515	-0.023	nPVI-V	-0.076	0.484

- Linear mixed-effects modeling was conducted, setting the **TD score** as the dependent variable.

4. Results & Discussion

- TD scores were higher for faster speech (**Fig. 3**).
- TD scores were also higher for speech with shorter and fewer pauses.
- The effect of PD appeared to be moderated when speech had higher Var-C (**Fig. 4**).
- DR seemed to negatively affect TD scores only when speech had higher Var-C (**Fig. 5**).

Fig. 3

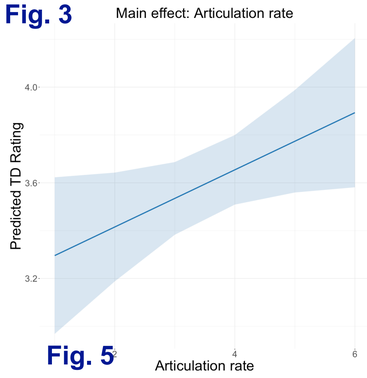


Fig. 1

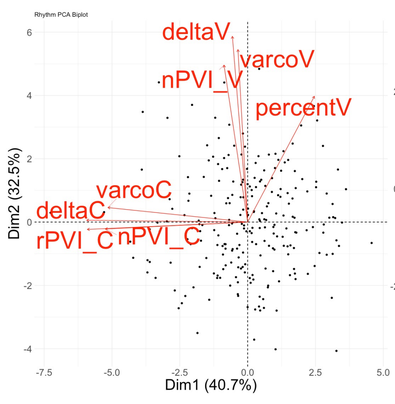


Fig. 2

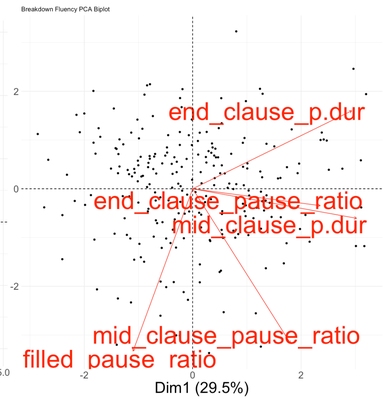


Fig. 4

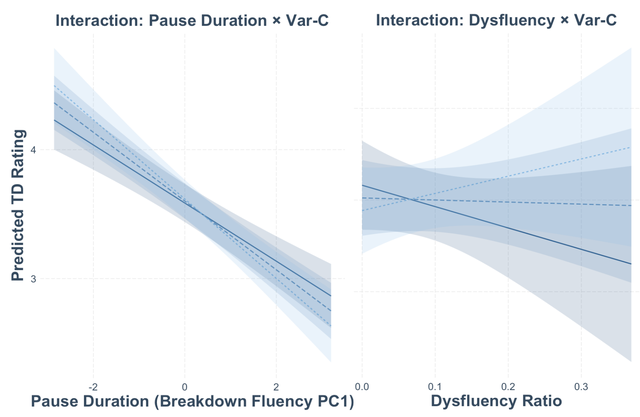


Fig. 5

	est	CI	*p*
(intercept)	3.18	2.72 −3.65	<.001*
articulation rate (AR)	**0.12**	**0.00 −0.24**	**<.05***
pause duration (PD)	**-0.27**	**-0.32 −-0.21**	**<.001***
pause frequency (PF)	**0.14**	**0.08 −0.20**	**<.001***
disfluency ratio (DR)	-0.09	-1.17 −0.99	.869
Var-C	0.06	-0.16 −0.28	.587
Var-V	0.17	-0.09 −0.43	.193
AR*Var-C	-0.01	-0.07 −0.05	.783
AR*Var-V	-0.05	-0.12 −0.02	.139
PD*Var-C	**0.02**	**-0.00 −0.05**	**.092**
PD*Var-V	0.01	-0.02 −0.04	.422
PF*Var-C	-0.02	-0.05 −0.01	.239
PF*Var-V	0.01	-0.02 −0.04	.512
DR*Var-C	**-0.47**	**-0.98 −0.05**	**.076**
DR*Var-V	-0.01	-0.64 −0.62	.976

References

[1] Kormos (2006). Speech production and second language acquisition. Mahwah, NJ: Lawrence Erlbaum Associates.
[2] Tavakoli, Nakatsuhara, & Hunter (2020). Aspects of fluency across assessed levels of speaking proficiency. The Modern Language Journal, 104, 169–191.
[3] McAuliffe et al. (2017). Montreal Forced Aligner: trainable text-speech alignment using Kaldi. In *Proc of the 18th ISCA Conf.*
[4] Matsuura, Suzuki et al. (2025). Gauging the validity of machine learning-based temporal feature annotation to measure fluency in speech automatically. Research Methods in Applied Linguistics, 4(1), 100177.