Reframe Your Thinking

2022 ASHA CONVENTION

Resilience REINVENTED

New Orleans • November 17-19

Virtual Library • November 10-28

Recognition of Aphasic Speech:
ASR Development and Analysis

Presentation by: Robert C. Gale[1] and Mikala Fleegle[2]
[1] Oregon Health & Science University, Portland, OR, USA
[2] Portland State University, Portland, OR, USA

Portland Allied Labs for Aphasia Technology (PALAT)

National Institute on Deafness and Other Communication Disorders

Disclosure

Financial: Research reported in this presentation was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award number R01DC015999.

Presentation Overview

1. ASR for Clinical Assessment of Anomia
2. Post-Stroke Speech Transcription Challenge
3. ASR Analysis Tool: PhonoLogic Viewer
   - Download: https://psst.study/phonologic/
4. Q&A and Discussion

ASR for Clinical Assessment

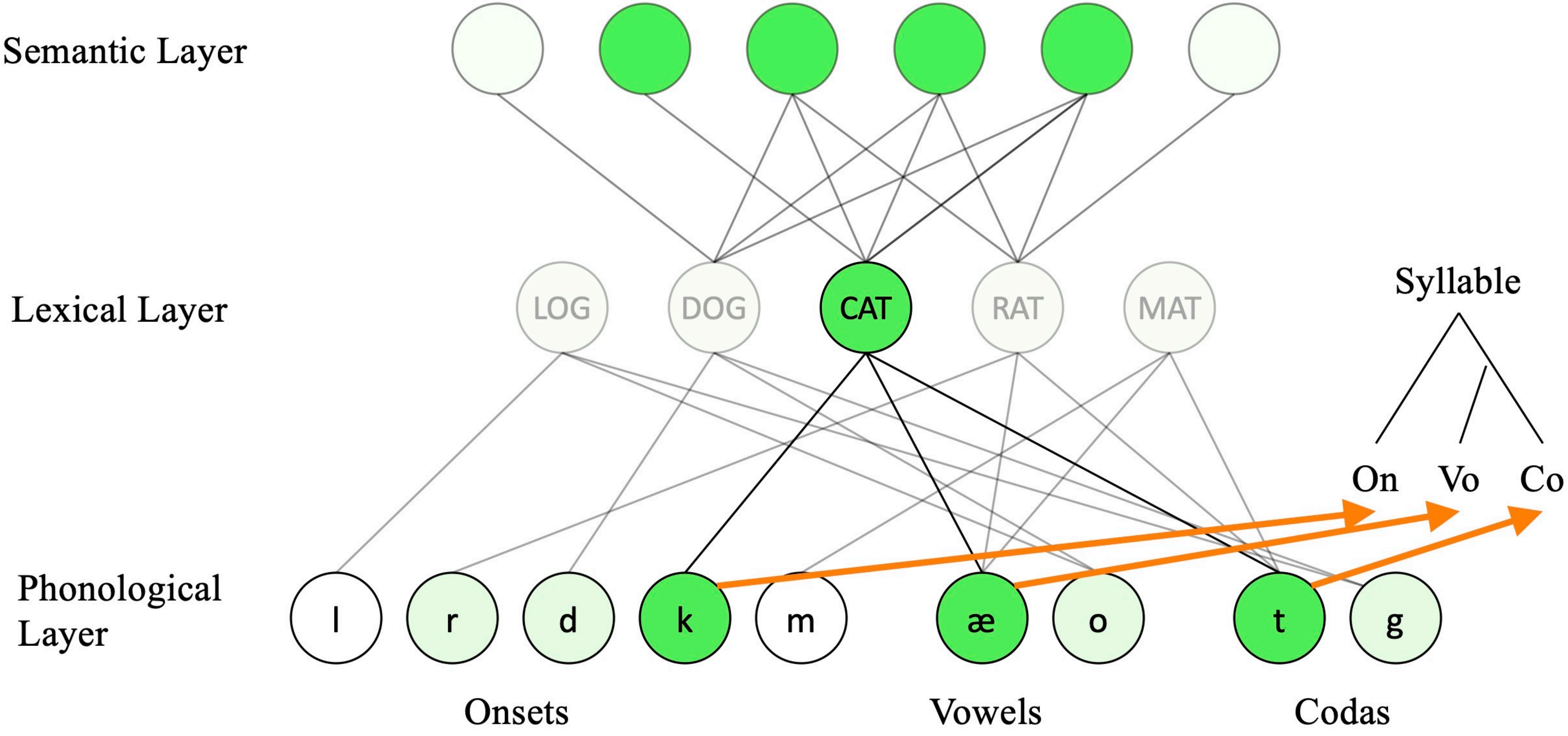
Who? people with aphasia

What? anomia

How? picture naming tests

Typical vs. Impaired Word Production

Dell's Model (Dell, 1986)

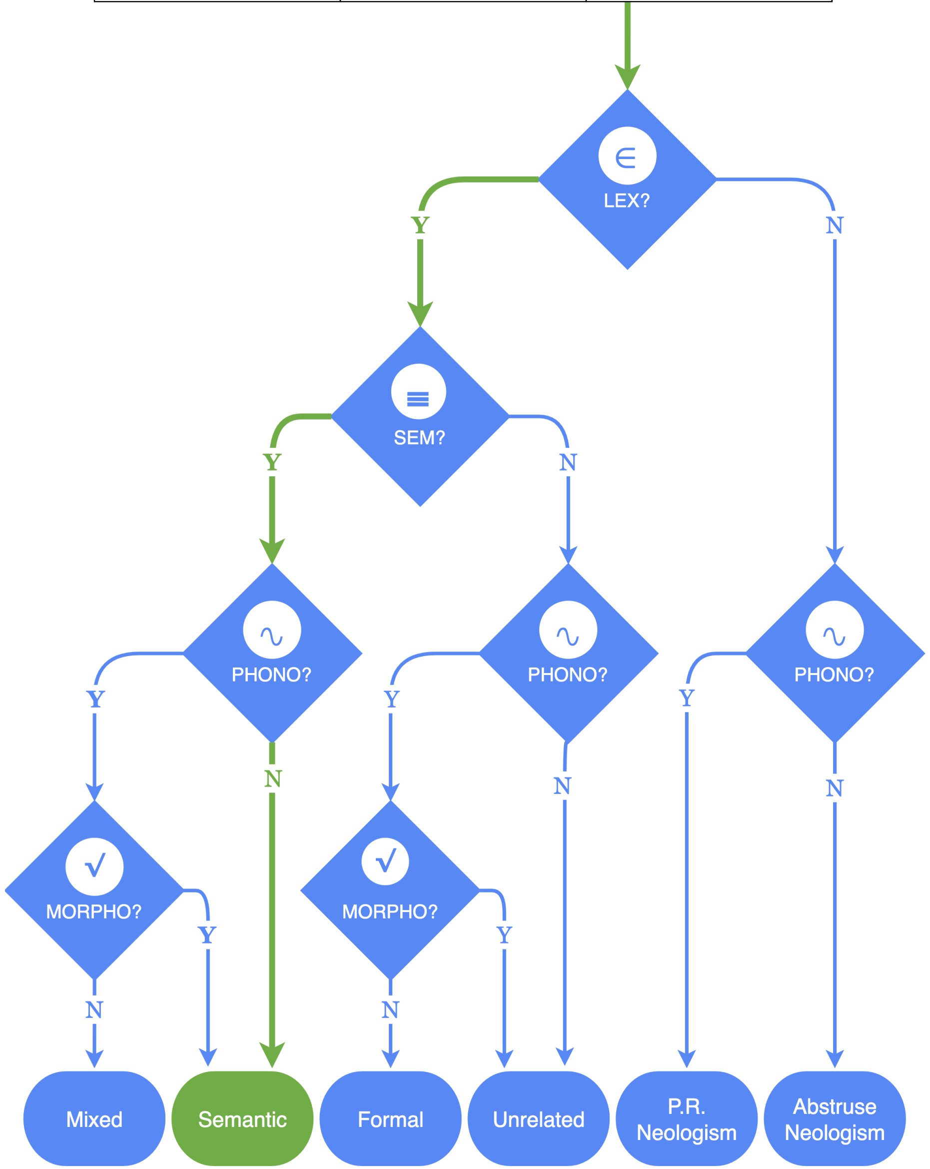


Anomia Assessment: Error Types and Analysis

	Paraphasia Type	Features			Example
		Lexical	Semantic	Phonological	
Lexical	Semantic	+	+	-	"dog" for "cat"
	Formal	+	-	+	"cot" for "cat"
	Mixed	+	+	+	"rat" for "cat"
	Unrelated	+	-	-	"mug" for "cat"
Non-Lexical	Neologism	-	n/a	+	"tat" for "cat"
	Abstruse Neologism	-	n/a	-	"vop" for "cat"

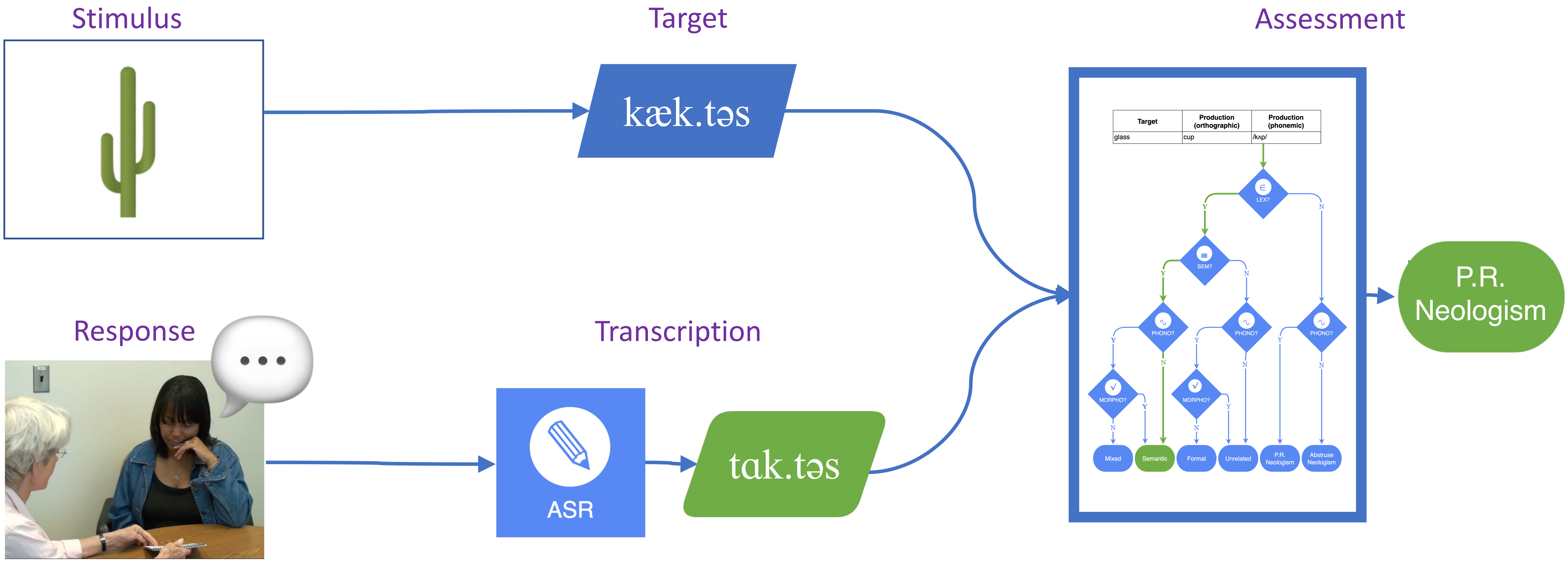
Anomia Assessment: The Value of Automation

Target	Production (orthographic)	Production (phonemic)
glass	cup	/kʌp/



Algorithmic Classification of Paraphasias aka “ParAlg” (Fergadiotis et al., 2016)

The Broader Vision:
Fully Automated Anomia Assessment



ASR: Commercial vs. Clinical

Commercial ASR:
/window/

Clinical ASR:
[bɪndoʊ]

Post-Stroke Speech Transcription (PSST) Challenge

LREC 2022
Marseille

PSST

(Gale et al., 2022)

The PSST Challenge

We provided

- A new dataset
  - Audio from English AphasiaBank (MacWhinney et al. 2011)
  - New phonemic transcripts
- A baseline phonemic ASR model
  - 26.4% phoneme error rate (PER)
  - 12.1% feature error rate (FER)

Challengers brought

- Clever new ideas
  - Several approaches to data augmentation
- An improvement on our baseline!
  - 20.0% phoneme error rate (PER)
  - 9.9% feature error rate (FER)

Gale et al. (2022) – https://aclanthology.org/2022.rapid-1.6/

PSST Speech Recognition Results

Yuan et al. (2022) →

Moëll/O’Regan. →
et al. (2022) .

Our baseline →

		Data (hours of audio)					ASR	
Model	**Arch**	**Pretrain**	**PSST**	**TIMIT**	**AphasiaBank**	**Other**	**FER**	**PER**
Y1	LARGE	60,000	2.8		33.3[U]		**9.9%**	**20.0%**
Y2	LARGE	60,000	2.8	3.9			10.3%	21.1%
Y3	LARGE	60,000	2.8		44.0[W]		10.4%	21.5%
Y4	LARGE	60,000	2.8			3.9[L]	10.6%	22.2%
Y5	LARGE	60,000	2.8				10.9%	22.3%
MO1	LARGE	960	2.8	1.1 [r]			11.3%	25.5%
MO2	LARGE	960	5.6 [p]				11.4%	25.1%
MO3	BASE	960	2.8	1.1 [r]			11.7%	26.3%
MO4	LARGE	960	5.6 [t]				11.7%	25.4%
MO5	LARGE	960	5.6 [p]	1.1 [r]			11.9%	26.0%
MO6	LARGE	960	2.8				12.0%	25.9%
MO7	BASE	960	5.6 [n]				12.0%	26.1%
PSST–A	BASE	960	2.8				12.1%	26.4%
Y6	LARGE	60,000	2.8			100[L]	12.5%	26.0%
Y7	LARGE	60,000	2.8			960[L]	16.7%	38.0%

[L] Librispeech, pseudo-labeled with G2P
[U] iteratively pseudo-labeled (unweighted)
[W] iteratively pseudo-labeled (weighted)
[p] with pitch-shifted variants
[t] with time-shifted variants
[n] with Gaussian noise augmentation
[r] RIR reverb applied

Evaluating an ASR

Word error rate (WER)

Orthographic ASR: $\frac{\#\,WORD\ ERRORS}{\#\,TARGET\ WORDS}$

Human:	a	house	🏠
ASR:	a	horse	
	✅	❌	$\frac{1}{2} = 50\%\ WER$

Phoneme Error Rate (PER)

Phonemic ASR: $\frac{\#\,PHONEME\ \ ERRORS}{\#\,TARGET\ PHONEMES}$

Human:	t	ɑ	k	t	ə	s	🌵
ASR:	d	ɑ	k	t	ə	s	
	❌	✅	✅	✅	✅	✅	$\frac{1}{6} = 17\%\ PER$

Further intuition: /taktəs/ → /daktəs / *should score better than* /taktəs/ → /oaktəs/

Phonological Features

p	=	<voiceless>	<bilabial>	<stop>
b	=	<voiced>	<bilabial>	<stop>
t	=	<voiceless>	<alveolar>	<stop>
d	=	<voiced>	<alveolar>	<stop>
k	=	<voiceless>	<velar>	<stop>
g	=	<voiced>	<velar>	<stop>

ARPAbet	IPA	consonantal	delayedrelease	continuant	sonorant	approximant	syllabic	tap	nasal	voice	spreadglottis	labial	round	labiodental	coronal	anterior	distributed	strident	lateral	dorsal
P	p	+	–	–	–	–	–	–	–	–	–	+	–	–	–	0	0	0	–	–
B	b	+	–	–	–	–	–	–	–	+	–	+	–	–	–	0	0	0	–	–
T	t	+	–	–	–	–	–	–	–	–	–	–	–	–	+	+	–	–	–	–
D	d	+	–	–	–	–	–	–	–	+	–	–	–	–	+	+	–	–	–	–
K	k	+	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	0	–	+
G	g	+	–	–	–	–	–	–	–	+	–	–	–	–	–	0	0	0	–	+

Distance between two phonemes

- Feature system: a table of distinctive features
  - Modified version of Hayes (2009)
  - 24 features x 40 phonemes
- Consider each phoneme as a set of features
- Error cost as a vector distance:

$$\text{Cost}(s, ʃ) = \left\| \overrightarrow{\text{s ʃ}} \right\| = \left\| \begin{bmatrix} +\text{cons} \\ +\text{delrel} \\ +\text{cont} \\ +\text{ant} \\ -\text{dist} \\ \ldots \\ -\text{voi} \end{bmatrix} - \begin{bmatrix} +\text{cons} \\ +\text{delrel} \\ +\text{cont} \\ -\text{ant} \\ +\text{dist} \\ \ldots \\ -\text{voi} \end{bmatrix} \right\| = \left\| \begin{matrix} 0 \\ 0 \\ 0 \\ -1 \\ 1 \\ \ldots \\ 0 \end{matrix} \right\| = 2 \text{ features apart}$$

Special considerations (don't worry too much about these)

- Values can be:
  - Present [+]
  - Absent [–] or
  - Not relevant [0]
- Diphthongs
  - Calculate as one phoneme or two?
  - Workaround, new values:
    - Absent-to-present [–+]
    - Present-to-absent [+–]

Cost	Feature Changes		
1	[–feature]	↔	[+feature]
0.75	[–feature]	↔	[+–feature]
	[–+feature]	↔	[+feature]
0.5	[–feature]	↔	[0feature]
	[–+feature]	↔	[+–feature]
	[0feature]	↔	[+feature]
0.25	[–feature]	↔	[–+feature]
	[–+feature]	↔	[0feature]
	[0feature]	↔	[+–feature]
	[+–feature]	↔	[+feature]
0	[–feature]	↔	[–feature]
	[–+feature]	↔	[–+feature]
	[0feature]	↔	[0feature]
	[+–feature]	↔	[+–feature]
	[+feature]	↔	[+feature]

Distance between two *transcripts*

- Similar to PanPhon (Mortensen, 2016)
- Find alignment with least error (Levenshtein, 1966)
- Insertions & deletions: ignore undefined features

Phoneme Error Rate (PER) ***vs.*** **Feature Error Rate (FER)**

Human:	l	æ	f			ɪ	n
ASR:	b	ɹ	ɑ	p	ɹ	ɪ	ŋ
	❌	❌	❌	❌	❌	✅	❌

$= \frac{6}{5} = 120\%$

Human:		l	æ	f		ɪ	n
ASR:	b	ɹ	ɑ	p	ɹ	ɪ	ŋ
	$\frac{22}{24}$	$\frac{4}{24}$	$\frac{2}{24}$	$\frac{3}{24}$	$\frac{23}{24}$	✅	$\frac{23}{24}$

$= \frac{58.5}{130} = 45\%$

Feature distance sounds very promising, but…

- Even when you understand the principles…
  - Unreasonable to estimate in your head
- Even when you're looking at the answer…
  - Difficult to explain why
- Cross-disciplinary: linguistics, computer science
- Cumbersome: dozens of features per phoneme, alignment

Don't fret, though…

PhonoLogic Viewer

an ASR analysis tool

Download:
https://psst.study/phonologic/

Questions?

References

Abel, S., Willmes, K., & Huber, W. (2007). Model-oriented naming therapy: Testing predictions of a connectionist model. *Aphasiology*, *21*(5), 411–447. https://doi.org/10.1080/02687030701192687

Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. *ArXiv, abs/2006.11477*.

Best, W., Greenwood, A., Grassly, J., Herbert, R., Hickin, J., & Howard, D. (2013). Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex*, *49*(9), 2345–2357. https://doi.org/10.1016/j.cortex.2013.01.005

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Edmonds, L. A., & Kiran, S. (2006). Effect of semantic naming treatment on crosslinguistic generalization in bilingual aphasia. *Journal of Speech Language and Hearing Research*, *49*(4), 729.

Fergadiotis, G., Gorman, K., and Bedrick, S. (2016). Algorithmic classification of five characteristic types of paraphasias. *American Journal of Speech-Language Pathology*, 25(4S):S776–S787, December.

Gale, R.C., Fleegle, M., Fergadiotis, G., Bedrick, S. (2022). The Post-Stroke Speech Transcription (PSST) Challenge. In *Proceedings of the RaPID-4,* LREC 2022, pages 62–70, Marseille, France. European Language Resources Association.

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9, pp. 133–177). Academic Press.

Graves, A., Fernández, S., Gomez, F.J., & Schmidhuber, J. (2006). Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. *Proceedings of the 23rd international conference on Machine learning*.

Hayes, B. (2009). Introductory Phonology. Wiley-Blackwell, Malde, MA.

Kendall, D. L., Rosenbek, J. C., Heilman, K. M., Conway, T., Klenberg, K., Gonzalez Rothi, L. J., & Nadeau, S. E. (2008). Phoneme-based rehabilitation of anomia in aphasia. *Brain and Language*, *105*(1), 1–17.

References (cont.)

Kertesz, A. (2007). *Western Aphasia Battery – R*. Grune & Stratton.

Levenshtein, Vladimir I. (February 1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady.* 10 (8): 707–710.

Mack, W. J., Freed, D. M., Williams, B. W., & Henderson, V. W. (1992). Boston Naming Test: Shortened versions for use in Alzheimer's disease. *Journal of Gerontology*, *47*(3), 154–158

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). Aphasiabank: Methods for studying discourse. *Aphasiology*, *25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893

Birger Moell, Jim O'Regan, Shivam Mehta, Ambika Kirkland, Harm Lameris, Joakim Gustafson, and Jonas Beskow. 2022. Speech Data Augmentation for Improving Phoneme Transcriptions of Aphasic Speech Using Wav2Vec 2.0 for the PSST Challenge. In *Proceedings of the RaPID-4,* LREC 2022, pages 62–70, Marseille, France. European Language Resources Association.

Mortensen, D.R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C., & Levin, L.S. (2016). PanPhon: A Resource for Mapping IPA Segments to Articulatory Feature Vectors. *COLING*.

Perez, M., Aldeneh, Z., & Provost, E. M. (2020). Aphasic Speech Recognition Using a Mixture of Speech Intelligibility Experts. *Interspeech 2020*, 4986–4990.

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Faseyitan, O., Brecher, A., Dell, G. S., & Coslett, H. B. (2009). Anterior temporal involvement in semantic word retrieval: Voxel-based lesion-symptom mapping evidence from aphasia. *Brain*.

Thompson, C. K. (2011). Northwestern Assessment of Verbs and Sentences. Evanston, IL.

Jiahong Yuan, Xingyu Cai, and Kenneth Church. 2022. Data Augmentation for the Post-Stroke Speech Transcription (PSST) Challenge: Sometimes Less Is More. In *Proceedings of the RaPID-4,* LREC 2022, pages 71–79, Marseille, France. European Language Resources Association.

THANK YOU FOR ATTENDING