



aphaDIGITAL – Avatar-based digital speech therapy solution for aphasia patients: first evaluation

Eugenia Rykova^{1,2}, Mathias Walther¹, Elisabeth Zeuner³

¹ Technical University of Applied Sciences Wildau

² University of Eastern Finland

³ Martin-Luther-Universität Halle-Wittenberg

eugeniia.rykova@th-wildau.de

Aphasia

Aphasia facts

- Aphasia is an acquired language disorder due to focal brain injury (a stroke in 80% cases)
- Annually, 270.000 people suffer a stroke in Germany, and 25.000 people - in Finland
- 30% of all stroke survivors experience initial aphasia, chronic deficits - in 20% of cases
- Predominant age group is 50-59 y.o.
- Generally affects more men than women (up to three times more)
- Significant benefits of high intensity and duration of speech and language therapy (SLT)

Project Motivation:

- patients with aphasia do not have access to sufficient SLT
- efficiency of digital therapy solutions
- absence of apps with speech analysis and individual verbal feedback/avatar

References



Johann Augustin Philipp Gesner describes *speech amnesia* (1770)

Armand Trousseau coins the term *aphasia* („speechlessness“) and gives a series of corresponding lectures (1861)

Pierre Paul Broca describes *aphémie* („loss of articulated speech“) due to a lesion to the left frontal lobe -> *Broca's area/aphasia* (1864)

Carl Wernicke describes *sensory aphasia* (fluent but nonsense speech) due to lesions to the left temporal lobe -> *Wernicke's zone/aphasia* (1874)

(Neuro)psychology, linguistics, and neuroimaging contribute to a new understanding of the functional neuroanatomy of language (XX century)

Raise of aphasia awareness due to Bruce Willis' diagnosis (2022)

Speech Recognition Evaluation

- 40+ ASR systems: four commercial/36+ open-source, two of which based on phone recognition
- Test corpora (German):
 - PD - PHONDAT 2 (read-aloud speech)
 - CI - read-aloud speech of normal hearing speakers and speakers with cochlear implants
 - ALC - Alcohol Language Corpus (read-aloud and spontaneous speech; intoxicated/sober)
 - Stark - spontaneous speech of a female speaker with aphasia (AphasiaBank)
 - Besuch - different speech tasks performed by eight male speakers with aphasia (YouTube)

Word Information Lost				Word Error Rate				Character Error Rate			
all corpora		w/o PD, CI, Sportplatz 27		all corpora		w/o PD, CI, Sportplatz 27		all corpora		w/o PD, CI, Sportplatz 27	
model	av. WIL rank	model	av. WIL rank	model	av. WER rank	model	av. WER rank	model	av. CER rank	model	av. CER rank
google	0,46	7,3	google	0,62	8,4	google	0,38	6,1	HF/mfleck	0,50	5,4
fraunhofer	0,49	8,8	HF/mfleck	0,64	6,8	fraunhofer	0,42	8,7	HF/ jonatas-grosman2	0,53	6,2
HF/mfleck	0,54	8,8	HF/ jonatas-grosman1	0,65	5,6	HF/mfleck	0,46	8,7	HF/ jonatas-grosman1	0,54	6,4
HF/ jonatas-grosman2	0,55	8,6	HF/ jonatas-grosman2	0,66	7,4	IBM watson	0,48	11,9	google	0,54	7,6
IBM watson	0,56	11,9	HF/ oliverguhr2	0,67	10,8	HF/ jonatas-grosman2	0,50	8,9	HF/ oliverguhr2	0,55	11,6
HF/ oliverguhr2	0,56	13,6	ims_0	0,68	9,0	HF/ jsnfly	0,52	13,9	HF/ maxidl	0,59	13,8
HF/ jonatas-grosman1	0,57	10,4	fraunhofer	0,69	12,4	HF/ jonatas-grosman1	0,52	11,9	ims_0	0,59	12,8
ims_0	0,59	13,1	ims_70	0,69	11,6	HF/ oliverguhr2	0,53	13,5	fraunhofer	0,59	12,2
HF/ jsnfly	0,59	13,4	vosk_06	0,71	13,2	HF/ facebook	0,53	15,8	HF/ facebook	0,60	13,4
vosk_06	0,59	13,8	vosk_021	0,71	12,8	ims_0	0,54	15,3	HF/ oliverguhr1	0,60	14,8
vosk_021	0,59	13,6	HF/ jsnfly	0,71	12,8	HF/ oliverguhr1	0,54	14,4	IBM watson	0,60	14,8
HF/ maxidl	0,60	16,0	HF/ facebook	0,71	13,4	HF/ marcel	0,54	16,8	HF/ jsnfly	0,60	13,8
HF/ oliverguhr1	0,60	14,9	IBM watson	0,71	13,8	HF/ maxidl	0,55	14,0	HF/ marcel	0,62	16,4
e ml	0,62	22,7	e ml	0,81	27,2	e ml	0,63	23,6	e ml	0,86	28,8

Avatar Evaluation

Typ 1: NUDELN, REIS, KARTOFFELN, BROT

Typ 2: NUDELN, REIS, KARTOFFELN, BROT

Typ 3: General rating: Which school grade do you give the digital Therapy-assistance? (1-6)

Typ 4: Voice, Responsiveness, Eye contact, Support, Motivation

Experiment

- Introduction & Welcome
- Five picture-naming tasks with feedback
- Avatar evaluation scale (quantitative analysis)
- Questionnaire (qualitative analysis)

Participants

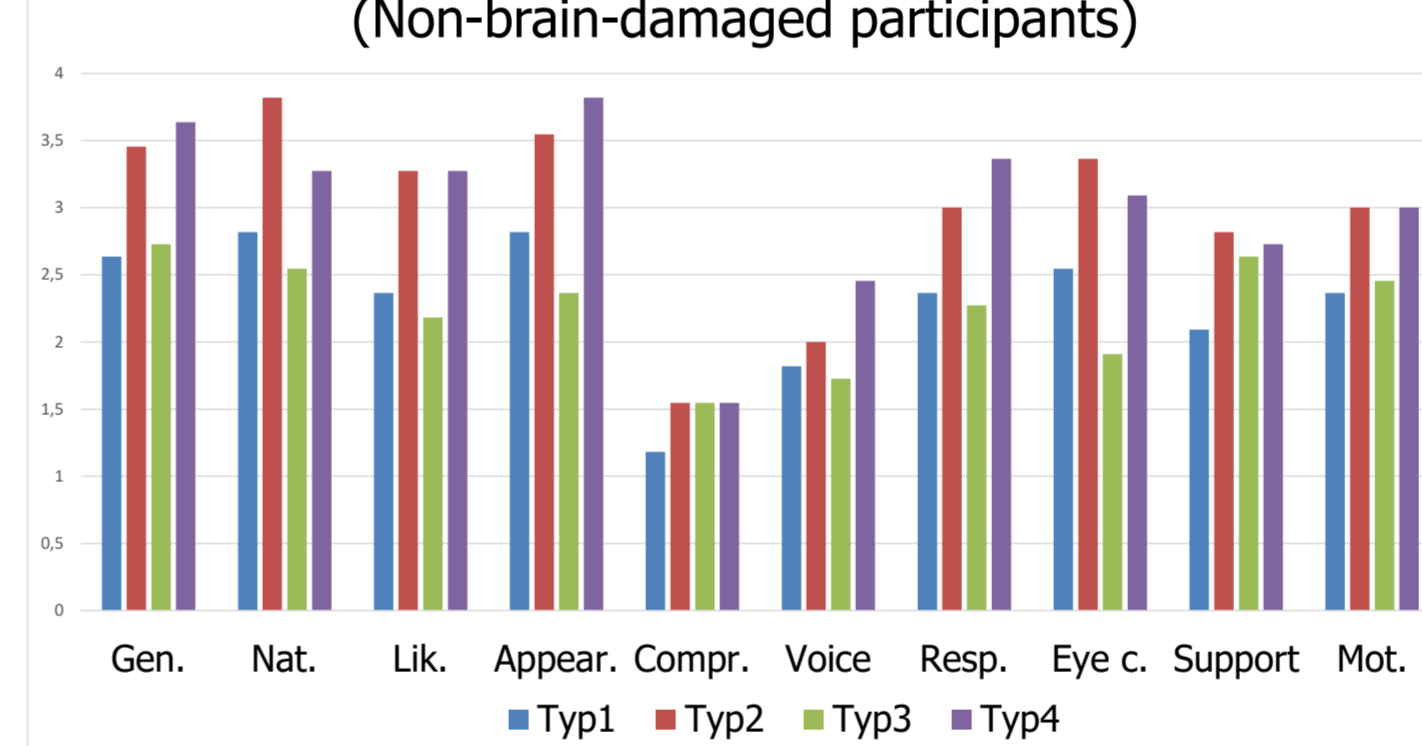
- 11 non-brain-damaged adults (6 women), 55-86 y.o. (M = 68)
- Four adults with aphasia (1 woman), 57-62 y.o. (M = 59)

Results

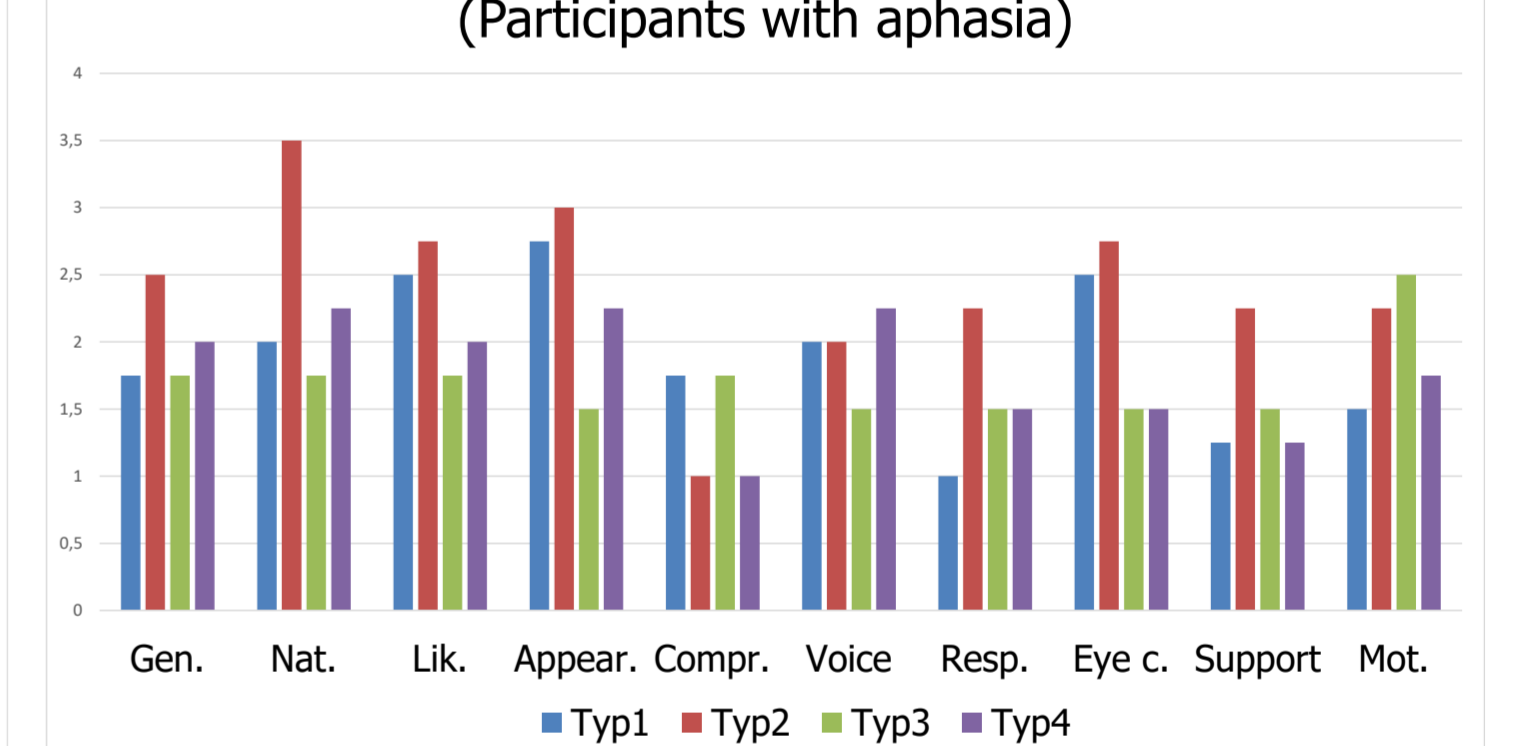
	Gen.	Nat.	Lik.	Appear.	Compr.	Voice	Resp.	Eye c.	Support	Mot.
Gen.	1,00									
Nat.	0,75	1,00								
Lik.	0,79	0,67	1,00							
Appear.	0,77	0,79	0,73	1,00						
Compr.	0,20	0,13	0,24	0,13	1,00					
Voice	0,66	0,66	0,54	0,61	0,31	1,00				
Resp.	0,79	0,86	0,71	0,77	0,27	0,74	1,00			
Eye c.	0,61	0,70	0,65	0,71	0,28	0,59	0,68	1,00		
Support	0,30	0,40	0,44	0,43	0,23	0,43	0,42	0,42	1,00	
Mot.	0,83	0,72	0,69	0,74	0,14	0,64	0,72	0,61	0,49	1,00

	Gen.	Nat.	Lik.	Appear.	Compr.	Voice	Resp.	Eye c.	Support	Mot.
Gen.	1,00									
Nat.	0,84	1,00								
Lik.	0,57	0,53	1,00							
Appear.	0,64	0,56	0,80	1,00						
Compr.	-0,24	-0,43	0,11	-0,14	1,00					
Voice	0,53	0,44	0,63	0,54	0,28	1,00				
Resp.	0,26	0,22	0,34	0,06	-0,04	0,08	1,00			
Eye c.	0,49	0,67	0,76	0,69	-0,02	0,35	0,05	1,00		
Support	0,22	0,26	0,51	0,15	0,23	0,30	0,72	0,22	1,00	
Mot.	0,30	0,29	0,21	0,24	0,00	0,00	0,00	0,24	0,25	1,00

Average ratings according to the avatar type (Non-brain-damaged participants)



Average ratings according to the avatar type (Participants with aphasia)



- Take home:
- slow speech desired
 - gender & age do not matter
 - prominent facial expressions desired

Error Analysis and Feedback

Picture to name: Target word: Apfel (Apple), Semantic Group: Frucht (Fruit) Semantic Class: Lebensmittel (Food)

Speech input 1: ASR → Afel (not a word) Phonemic/phonetic error (CER = 0.2) → CER threshold → Pronunciation match, Word match, Group match, Class match → Feedback: You are close! Look at the written word and read it carefully: „APFEL“.

Speech input 2: ASR → Birne (Pear) Semantic error (CER = 1) → GermaNet → Pronunciation match, Word match, Group match, Class match → Feedback: Not quite right. It is a different type of "Frucht". Please try again.

Speech input 3: ASR → Pirne (not a word) Phonemic/phonetic + semantic error (CER = 1) → GermaNet → Pronunciation match, Word match, Closest match: Birne, Group match, Class match → Feedback: Did you mean "Birne"? Not quite right. It is a different type of "Frucht". Please try again.

Speech input 4: ASR → Milch (Milk) Semantic error (CER = 1) → GermaNet → Pronunciation match, Word match, Group match, Class match → Feedback: It is a different type of "Lebensmittel". Please try again.

Speech input 5: ASR → Ball (Ball) Semantic error (CER = 1) → GermaNet → Pronunciation match, Word match, Group match, Class match → Feedback: I am sorry, "Ball" is not correct. Please try again.