Building a Spoken Dialogue System for Supporting Blind People in Accessing Mathematical Expressions

Pier Felice Balestrucci pierfelice.balestrucci@unito.it
Alessandro Mazzei alessandro.mazzei@unito.it
Luca Anselma luca.anselma@unito.it
Cristian Bernareggi cristian.bernareggi@gmail.com

Università degli Studi di Torino
Turin, Italy

1. Problem

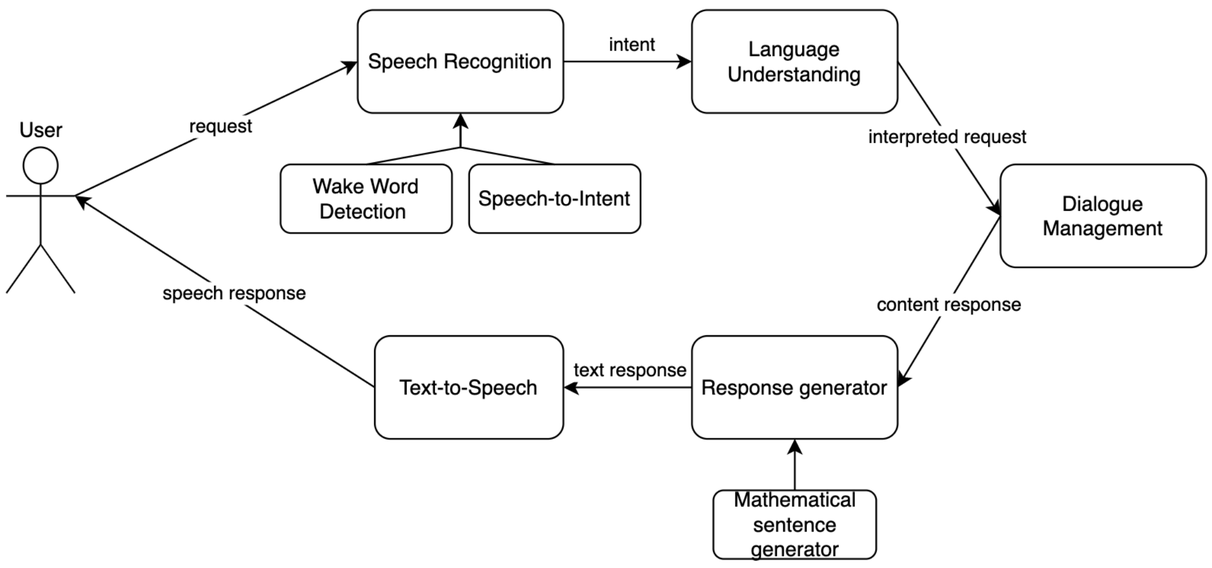
Visually impaired people (VIP) access visual information by listening to screen readers. Mathematical formulas are usually written in LaTeX to make them accessible:

- Knowledge of LaTeX is necessary
- It is verbose and heavy to listen to
- LaTeX can make symbols that are close, like fractions, appear very distant.

2. Objectives

- Designing a spoken dialogue system (SDS) to enhance access to visual information for VIP
- Investigating whether a SDS is accessible and usable for accessing mathematical formulas
- Enable the SDS to navigate mathematical formulas, allowing users to query specific parts of the expression or repeat sections

3. Spoken Dialogue System Architecture



4. From LaTeX to Enhanced Content Mathematics Mark-up Language (CMML)

$$\$y = f(x)\$$$

LatexML

```
<m:apply>
   <m:eq/>
   <m:ci>y</m:ci>
   <m:apply>
      <m:times/>
      <m:ci>f</m:ci>
      <m:ci>x</m:ci>
   </m:apply>
</m:apply>
```

Post Processing

```
<apply>
   <eq/>
   <ci>y</ci>
   <apply>
      <ci>f</ci>
      <ci>x</ci>
   </apply>
</apply>
```

5. Building Mathematical Sentences with NLG

- Document Planning → CMML
- Sentence Planning → Prototypical sentence
- Realization → SimpleNLG

6. From CCML to Mathematical Expressions

Algebraic, Arithmetic and Set operators

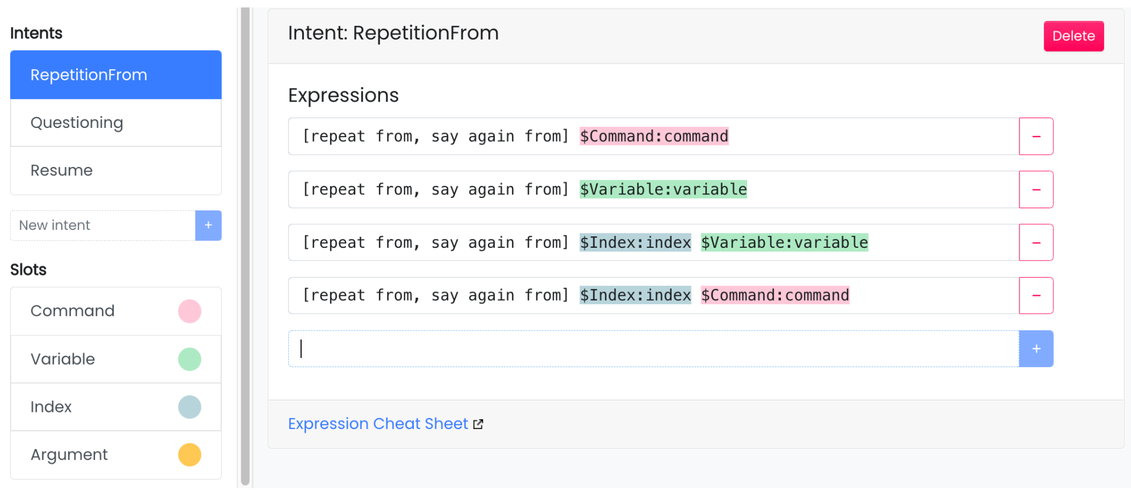
Symbol	Operator	English Form
+	plus	plus
−	minus	minus
/	divide	over
*	times	times
$[\ldots]^{[..]}$	power	to
\	settdiff	minus
×	cartesianproduct	cross
∩	intersect	the intersection of
∪	union	the union of ... and ...

7. Speech Synthesis

- **SSML (Speech Synthesis Markup Language)**
- **2 Speech Engines:** eSpeak & Amazon Polly

8. Speech Recognition and Language Understanding

- **Porcupine**: a pretrained model trained to recognized the wake-up word “Hey System!”
- **Rhino**: a pretrained model trained to convert the speech input to intents in a very specified domain



9. Dialogue Management

1. Parse the request
2. If the intent is "Questioning":
3. Search for the operator in the request
4. Match the request with the operator arguments
5. If the match is successful:
6. Answer the question
7. Otherwise, utter an apology message
8. If the intent is "RepetitionFrom":
9. Search within the sentence uttered what the user wants to hear repeated
10. If the search is successful:
11. Repeat from that point
12. Otherwise, utter an apology message
13. If the intent is "Resume":
14. Resume the synthesis of the mathematical expression

10. Experiment 1

Language: English | **Method:** Google Form with audio files

Users: 2 VIP | **10 Instances** (5 vs 5 - eSpeak & Amazon Polly)

User	Metrics	Tot. (25)	Easy (10)	Difficult (15)
1	EM	0.92	1.00	0.87
	SPICE	0.98	1.00	0.97
2	EM	1.00	1.00	1.00
	SPICE	1.00	1.00	1.00
avg	EM	0.96	1.00	0.93
	SPICE	0.99	1.00	0.99

The averaged SPICE and EM means and standard deviations for Experiment 1

11. Experiment 2

Language: English | **Method:** interaction with the SDS

Users 5 VIP | **6 Instances** (Amazon Polly)

User	Metrics	Tot. (6)	Easy (3)	Difficult (3)
1	EM	0.50	0.67	0.33
	SPICE	0.86	0.89	0.87
2	EM	1.00	1.00	1.00
	SPICE	1.00	1.00	1.00
3	EM	0.33	0.33	0.67
	SPICE	0.66	0.98	0.82
4	EM	0.33	0.00	0.67
	SPICE	0.80	0.95	0.89
avg	EM	0.54	0.50	0.67
	SPICE	0.83	0.95	0.89

The averaged SPICE and EM means and standard deviations for Experiment 2

Attribute	Score	Comparison to benchmark
Attractiveness	1.83	between 75% and 90%
Perspicuity	1.28	between 50% and 75%
Efficiency	1.33	between 50% and 75%
Dependability	1.33	between 50% and 75%
Stimulation	1.75	above 90%
Novelty	1.92	above 90%

User Experience Questionnaire Benchmark