1 Title

Good afternoon, my name is Tomáš Hanakovič and I'd like to present our ideas about how speech recognition can help visually impaired people to write mathematical formulas.

2 Introduction

Mathematical expression are very common, in many science domains, so it's desirable to pay attention how to create and publish them effectively.

No doubt that the fastest way is to write them manually, but these days we often need them in some kind of electronic form.

There are several editors available, some of them are visual and some of them not. But for both types we can discuss if they are suitable for disabled people. Visual solutions may be almost unusable because of complicated interface, in contrast the text based solution can be more accessible, but can have too complicated syntax rules to remember.

3 Motivation - multimodal approach

In practice we can see that application are restricted to one modality, mainly visual. We can ask: why this restriction?

Multimodal application offers more freedom – you can use mouse or voice as you wish. The same task can be done faster and much comfortable using right modality. For example navigation in formula can be done more faster using mouse, while entering new data is faster saying one sentence.

4 Dictation and mathematical formulas

Now let me sketch some problems, which we can have while dictating mathematical formulas.

5 Ambiguity of formula dictation

Let's start with simple example:

Consider sentence ***a i plus b j***. It has at least two interpretations, first one is sum two products: ***a times i and b times j***, second one is sum of indexed variables: ***a with index i and b with index j***. I don't know how it sounds to you, but my intuition tells me that the second one is more probable. But can I count on intuition? The answer is no, I can't.

So what's the formulas main feature that causes this ambiguity?

6 2D to 1D – linearisation

It is very important to realize that mathematical expressions have two dimensional nature – same looking elements can have different meaning because of their position and size – for example constant 2 versus index 2.

In contrast, sentences are in nature sequence of words, so they have only one dimension.

Right now we formulated the main problem – how to describe two-dimensional structure of mathematical expressions in simple one-dimensional sentence. And also, how to do it intuitively, but still preserve an exact interpretation! In fact, what we need is lossless reduction of two dimensions into one. We can call this process linearisation.

7 Analyse of formula structure

To make good linearisation we need to gather as much information as possible from the formula structure. Let us look at the structure closer.

At first we should identify structural patterns, like function, fraction, index and so on.

Also the recurrent structures are very common, for example fraction in fraction.

Very useful may be fact that we can logically split formula into smaller blocks, for example ***a squared plus two ab plus b squared*** can be seen as addition of three blocks. We call this splitting technique expansion.

8 Navigation

The second question related to formula structure is: how to mark specific atomic element in equation easily? Marking with mouse is reduced to single click, but what about voice? How to quickly navigate to that red b element in the displayed example?

We recommend hierarchical approach. It is simple and intuitive and quite fast.

The main idea is: consider formula as one big block. Expand it into smaller blocks. Select block that contains our element and expand it. Repeat with selections and expansions until block contains only single element, in our case that red b.

9 Hierarchical numbering – example

Here you can see the example: We want to select that „b“ element and correct it. First we select second block, which is first multiplication of sine and cosine. We expand this block to two sub-blocks – sine and cosine. We need select cosine - it is in the second block again. After expansion we got cosine value, which is that b we want to change, so now we are done with navigation.

As you can see this is very simple and quite fast way. All we need to remember in each step, is how current block is splitted and select appropriate sub-block for expansion. So it is very important to make split rules intuitive, so much the more for visually impaired people!

10 Application – key facts

These ideas were bring to life in application I wrote as part of my diploma thesis. That application was working presentation of techniques and approaches I've choose after problem analysis.

Only for fast preview – here is a short characteristics of what I wrote:

Application allows to enter mathematical formulas

it is full multimodal application – you work with independent but cooperative modalities – this means that each modality offers full set of functions, but using modalities cooperatively can improve speed and simplify the work.

Application is simple, easy to use thanks to simple user interface Moreover, it uses controls that are represented similar in visual and aural manner, for example menu. Voice commands offer synonyms and various sentence forms to be used.

Good orientation in formula is reached thanks to hierarchical approach.

Application uses clear colourful output in visual and aural part, which improves orientation in long readouts.

Also it has ability to be customized and configured at user will.

11 Resume

At the end let me resume some facts:

At first: Dictation of formulas can be ambiguous, but using good linearisation we can preserve formula structure and minimize word cost.

At second: Simple navigation and orientation in formula is very important for visually impaired people. Hierarchical approach leads to splitting formula to smaller blocks and selection/expansion steps, which is very intuitive and quite fast way of work.

At third: Voice command should allow synonyms and optional parameters to allow full flexibility and intuitiveness.

At fourth for output: adding colours to visual part and prosody to voice part can improve orientation in long outputs and attracts user attention.

12 Thank you for your attention!

And I thank YOU for your attention.