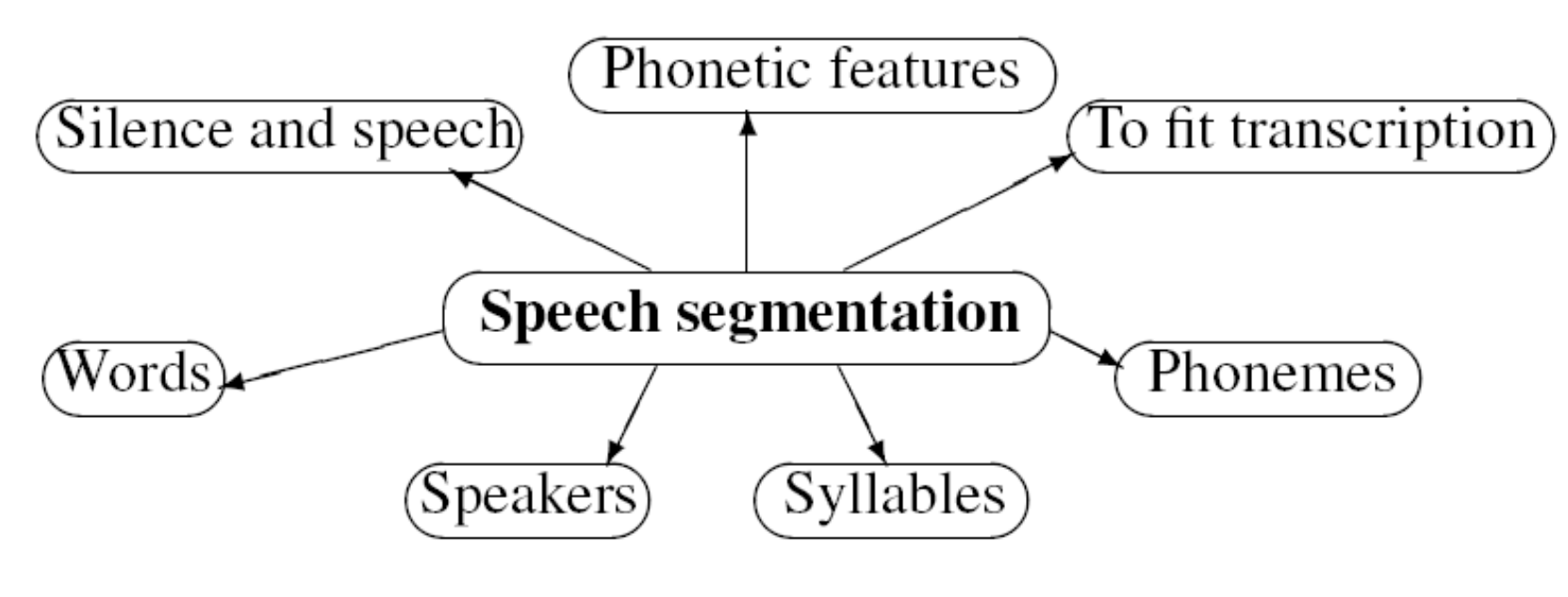
# Fuzzy Recall and Precision for Speech Segmentation

Bartosz Ziółko, Suresh Manandhar,
Richard C. Wilson

THE UNIVERSITY of York

# The types of speech segmentation



# Approaches to speech segmentation evaluation

- Counting number of insertions, deletions and substitutions
- Counting  the boundaries for which the deviation exceed thresholds (i.e. 35,70 ms)
- Value of tolerance is questionable
- Different inaccuracies treated in the same way
- Tolerances should be related to a length of an analysed phoneme

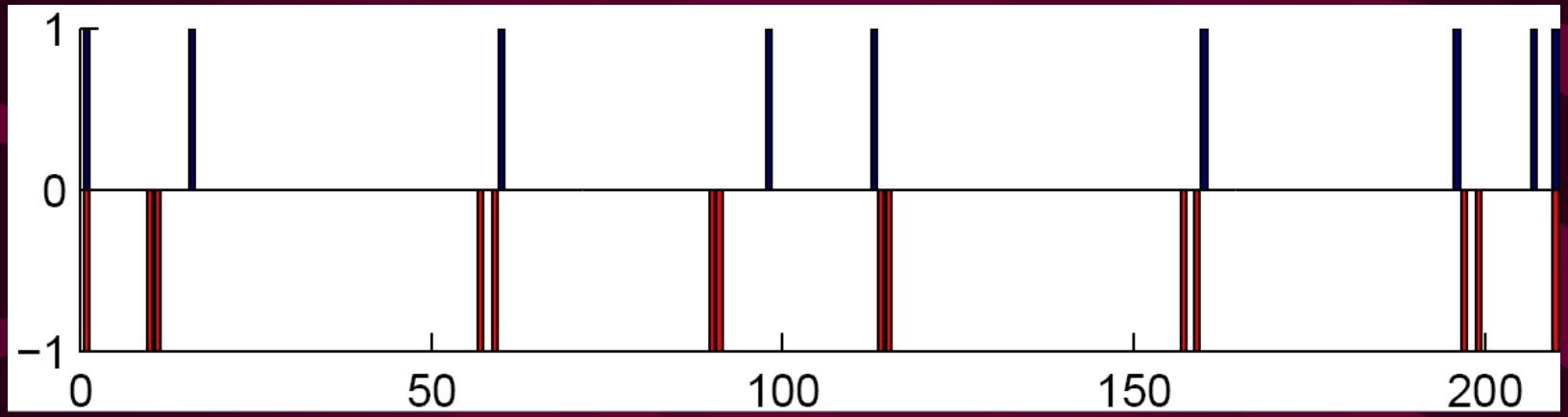
# Fuzzy logic

In a narrow sense fuzzy logic is considered a logical system aimed at providing a model for modes of human reasoning that are approximate rather than exact.

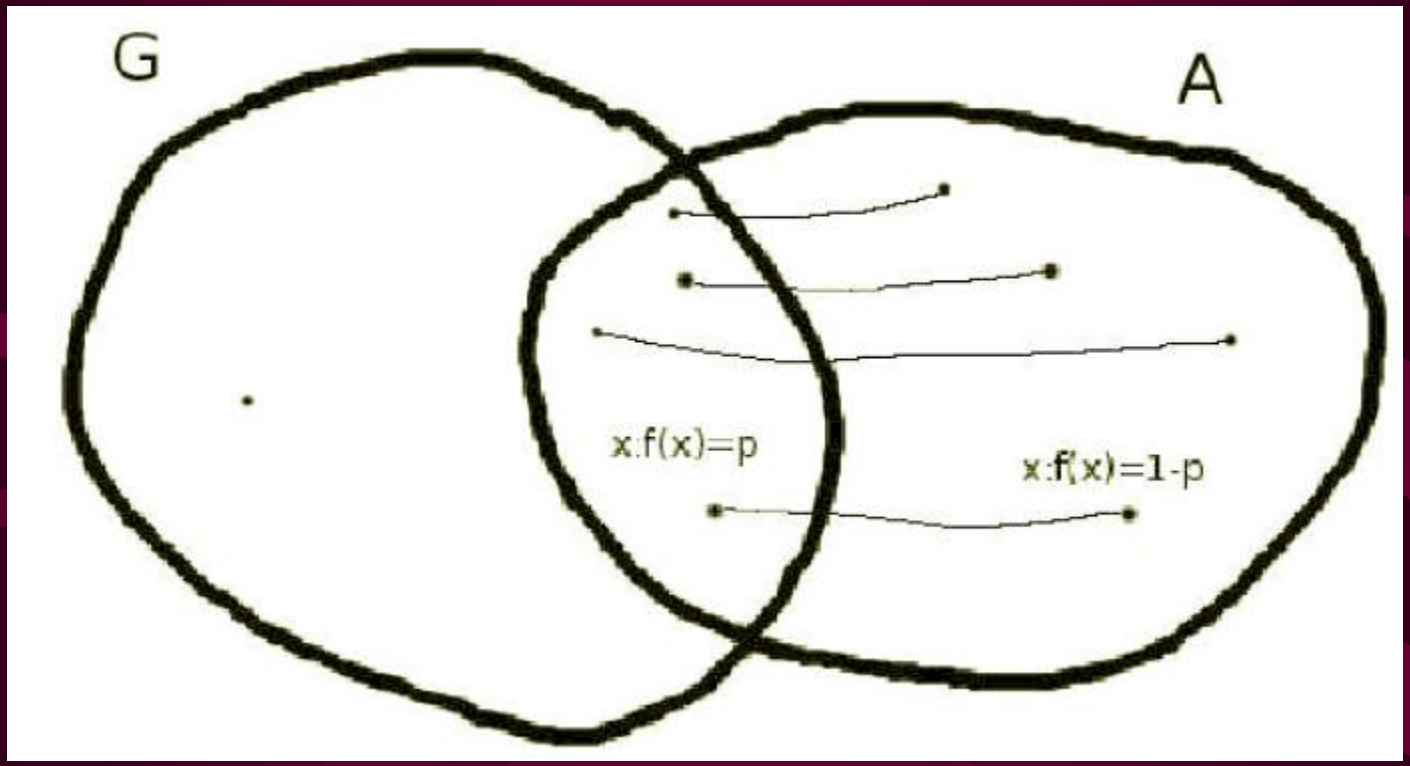
In a wider sense, it is treated as a fuzzy set theory of classes with unsharp boundaries

Detected boundaries may be shifted **more or less** with respect to a manual segmentation

# Example of phoneme segmentation



# The general scheme




G – correct boundaries, A- detected boundaries, precision is a comparision of number of relevant elements and the intersection, number of retrieved elements and intersection gives recall

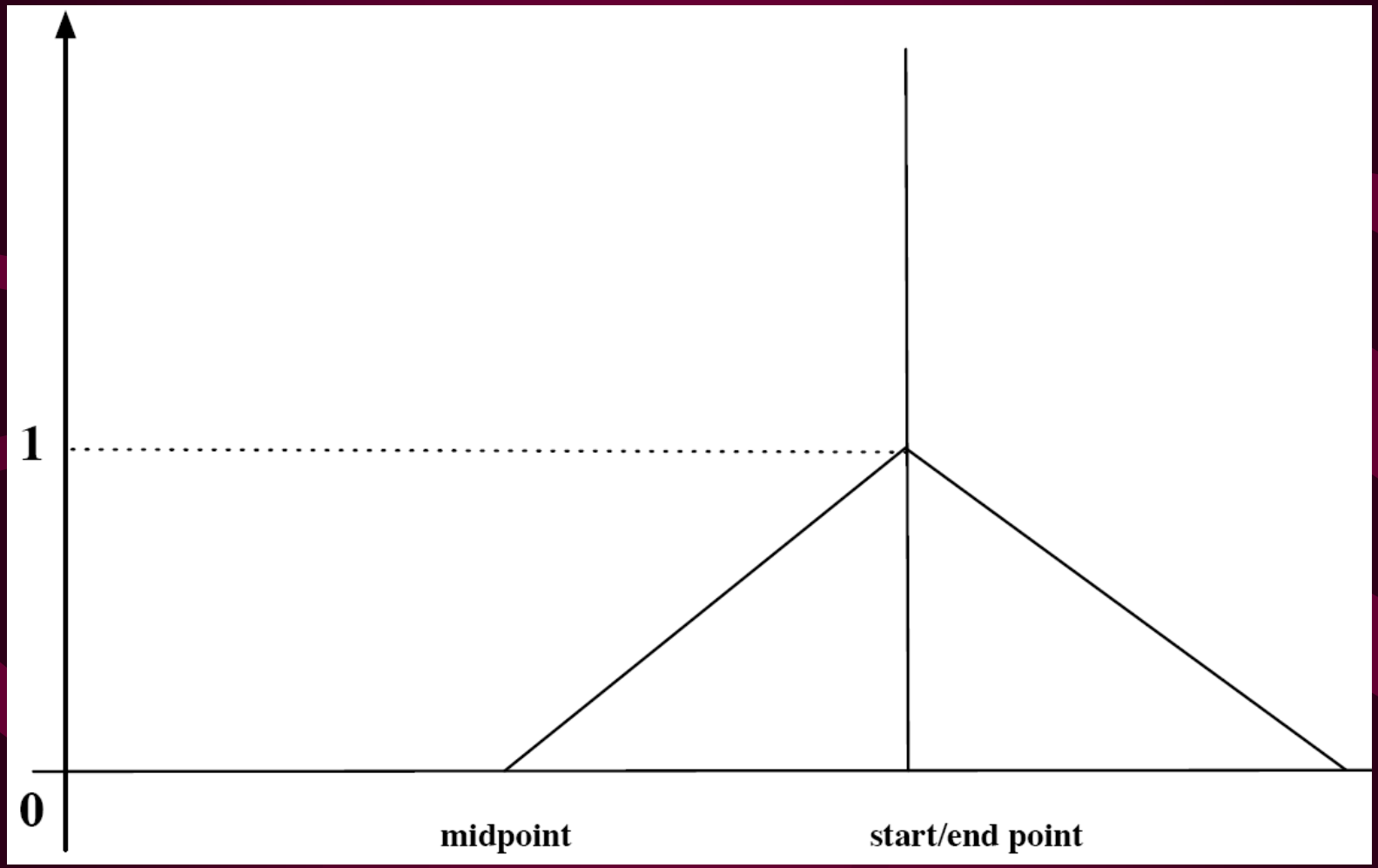
# Algorithm of evaluation

- Start with matching the closest detected and hand segmented boundaries.
- Calculate grades of being relevant and retrieved (next slide)
- Fuzzy precision
- Fuzzy recall

# Calculating membership function

- A hand segmented boundary not matched with any detected one $\in G$
- A detected boundary x not matched with any hand segmented $\in A$ and $f(x)=0$ (the last one)
- x is the hand segmented one so $\in A \in G$. $f(x)=1$
- Fuzzy case $f(x)=a–b/a$, $a$ is half of a length of the phoneme which boundary was detected, $b$ distance between hand segm and the detected one

# Fuzzy membership



# Comparision

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| beg | 9 | 56 | 89 | 113 | 156 | 196 | - |
| end | 10 | 58 | 90 | 114 | 158 | 198 | - |
| auto | 15 | 59 | 97 | 112 | 159 | 195 | 206 |
| fuzzy recall and precision | | | | | | | |
| f(x) | 0.78 | 0.93 | 0.36 | 0.91 | 0.95 | 0.95 | 0 |
| insertations and deletions without tolerance | | | | | | | |
| Ins(7) | X | X | X | X | X | X | X |
| Del(6) | X | X | X | X | X | X | - |
| with tolerance from 1 (5.8 ms) to 4 (23.2 ms) - same results | | | | | | | |
| Ins(3) | X | ✓ | X | ✓ | ✓ | ✓ | X |
| Del(2) | X | ✓ | X | ✓ | ✓ | ✓ | - |
| with tolerance 5 (29 ms) or 6 (34.8 ms) | | | | | | | |
| Ins(2) | ✓ | ✓ | X | ✓ | ✓ | ✓ | X |
| Del(1) | ✓ | ✓ | X | ✓ | ✓ | ✓ | - |
| with tolerance 7 (40.6 ms) or higher | | | | | | | |
| Ins(1) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | X |
| Del(0) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | - |

# Conclusions

- Precise evaluation method was described
- Correctness of audio segmentation is very rarely boolean
- F-score gives an option of favouring recall or precision

Thank You