

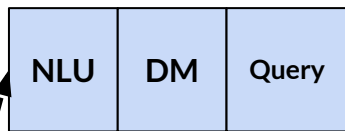
# InMind Movie Agent: Timing Results and Suggested Improvements

Vivian Tsai

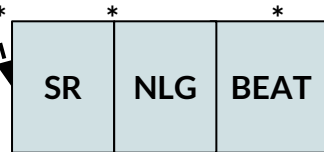
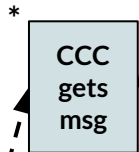
30 June 2017

# TurnExchange (with query)

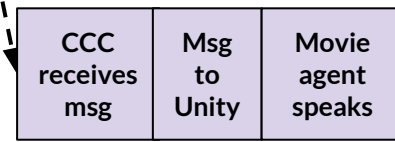
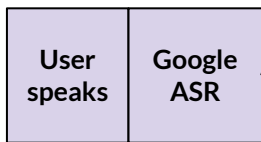
NLU/DM



MUF



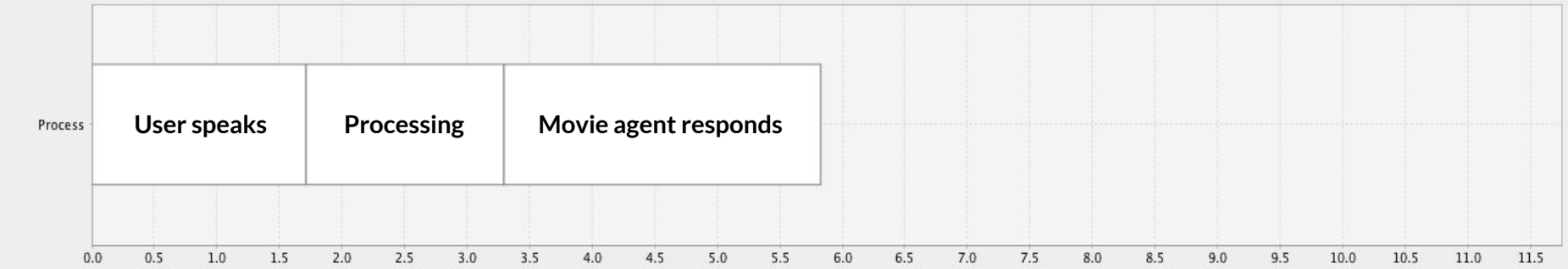
Phone



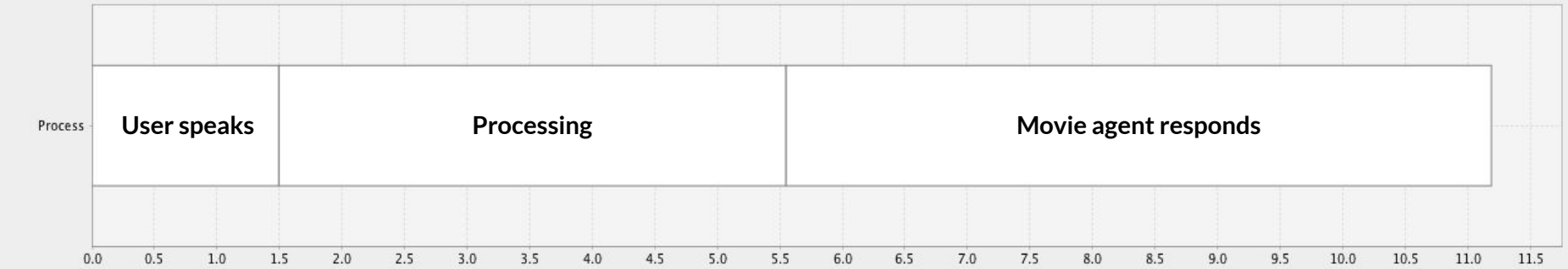
START

END

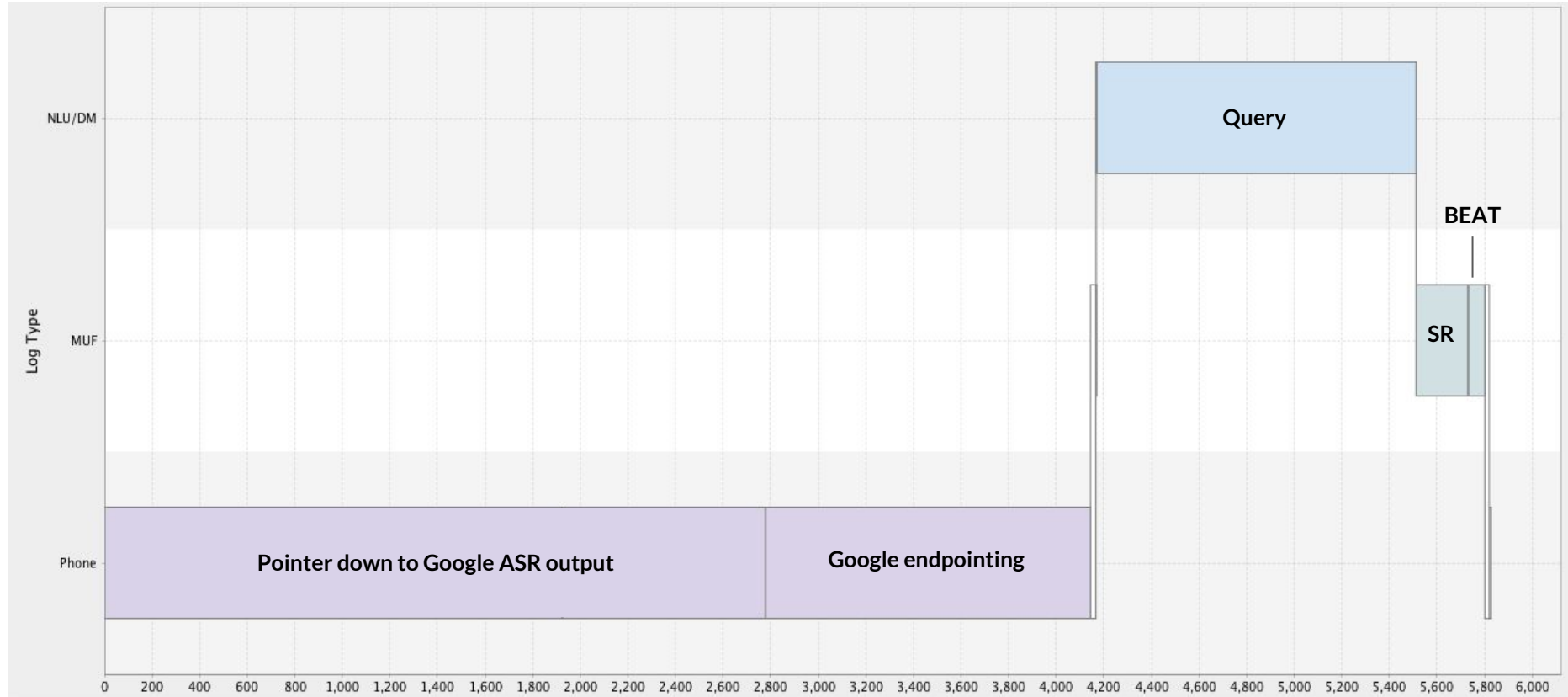
### Overall Timeline (no query)



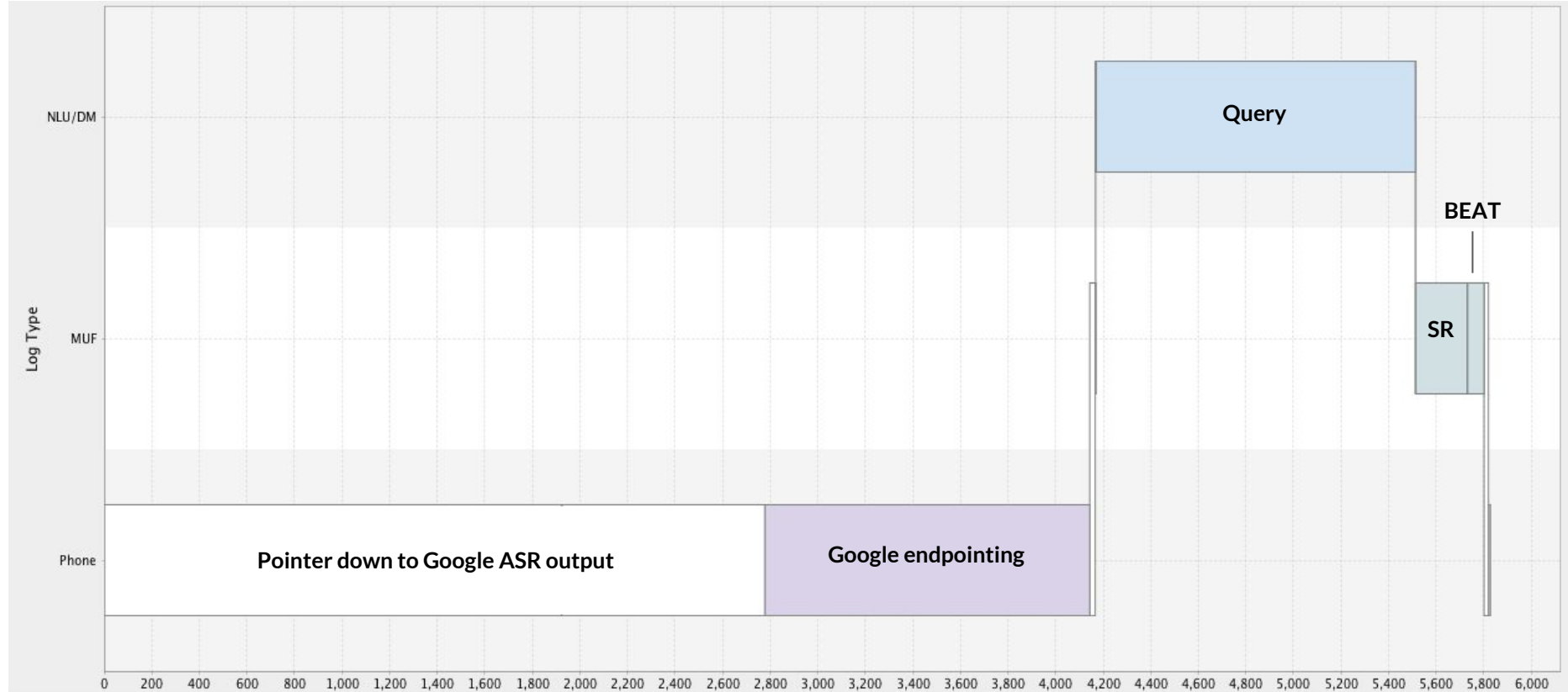
### Overall Timeline (with query)



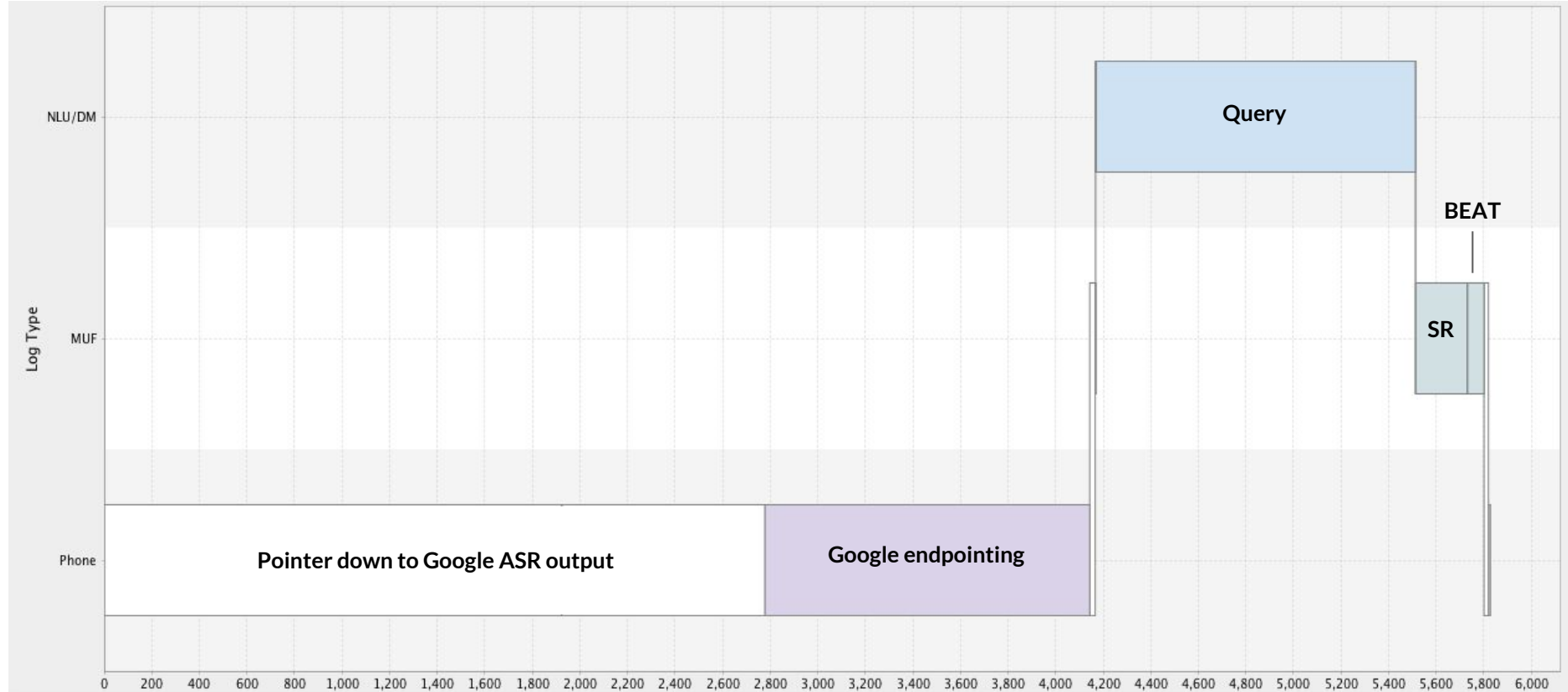
# Current Timeline (with Query)



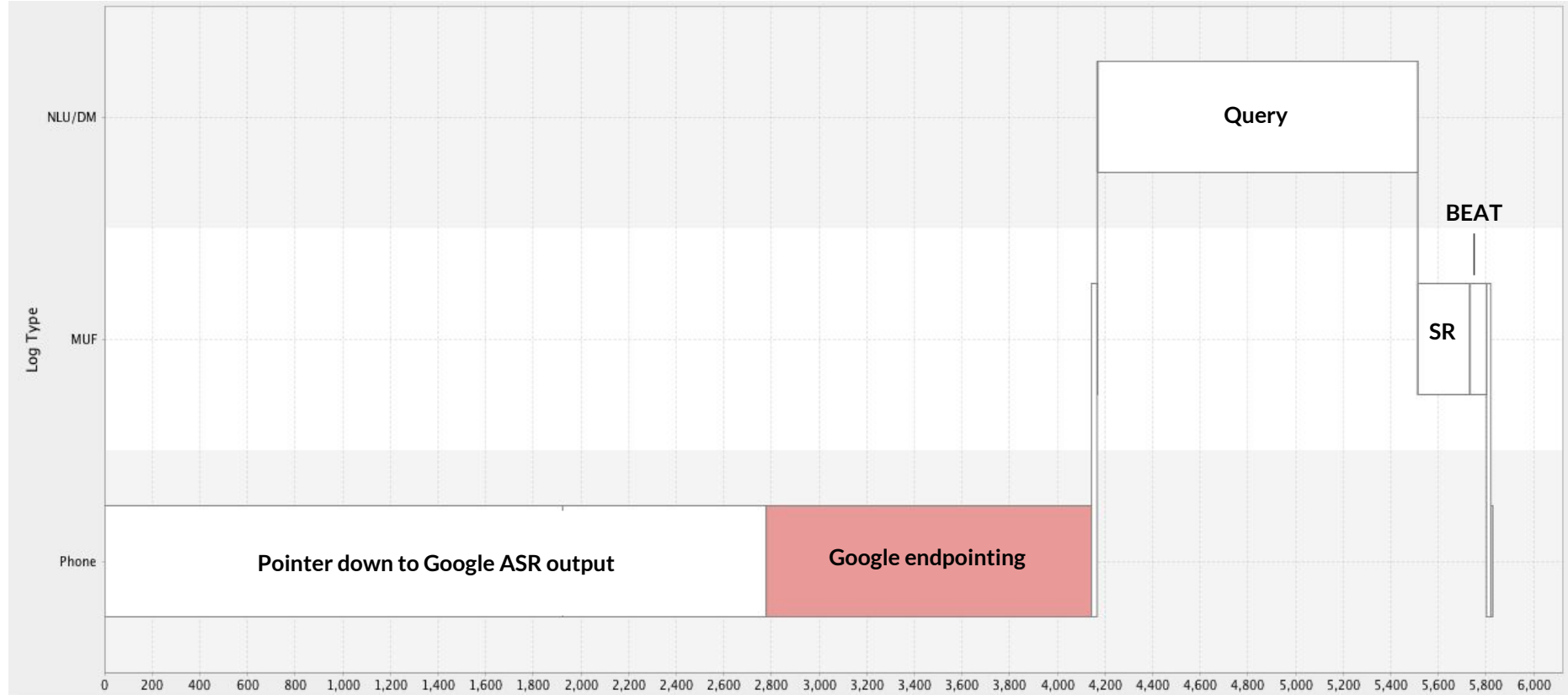
# Current Timeline (with Query)



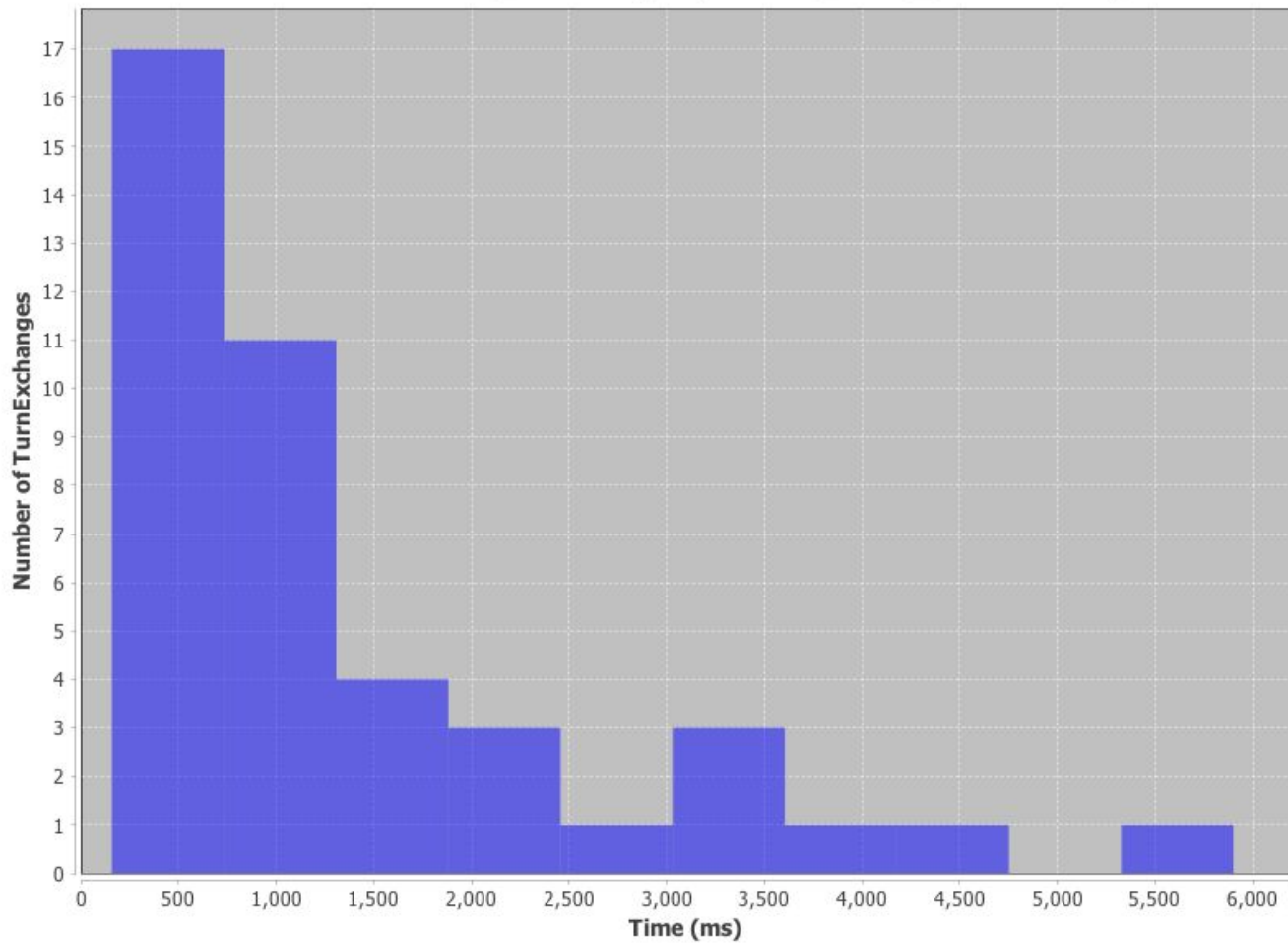
# Current Timeline (with Query)



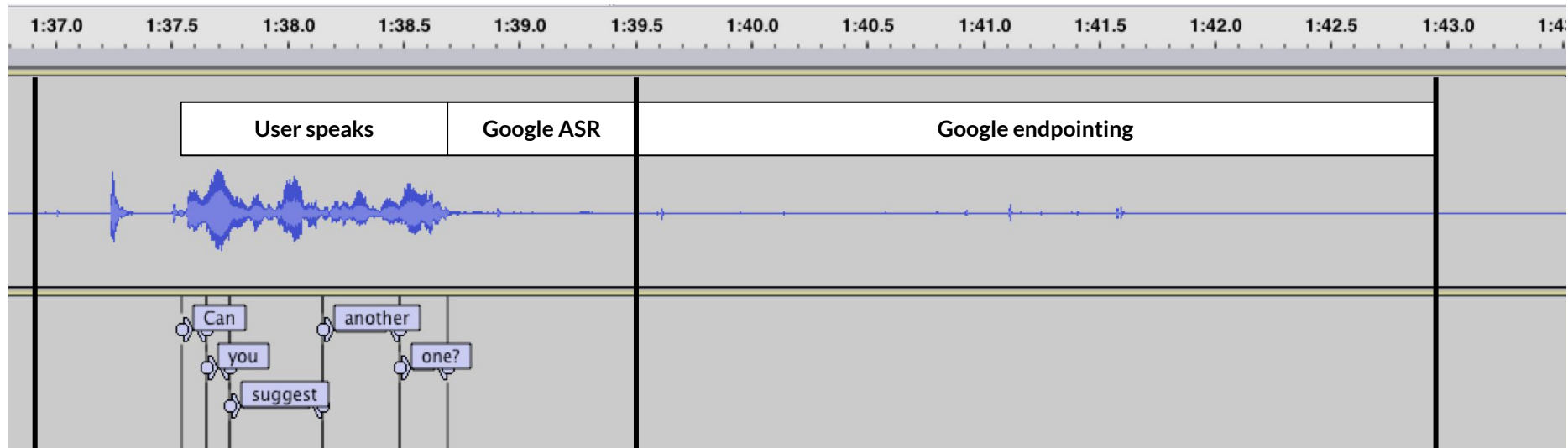
# Current Timeline (with Query)



**Time Between Phone\_ASROutput, Phone\_MUS\_SentMessage**



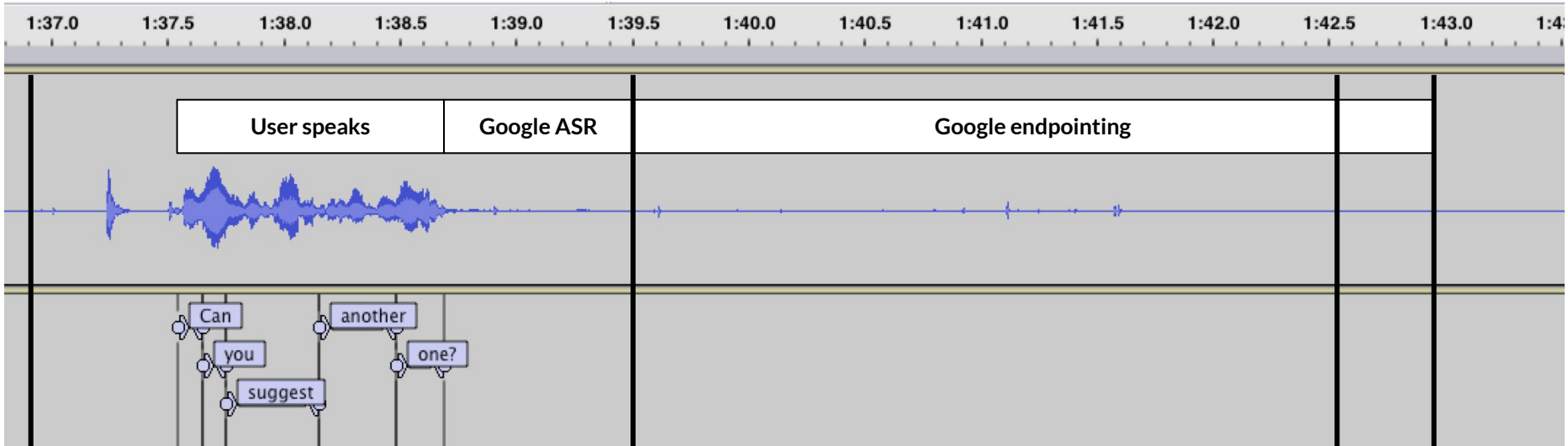




Pointer  
down

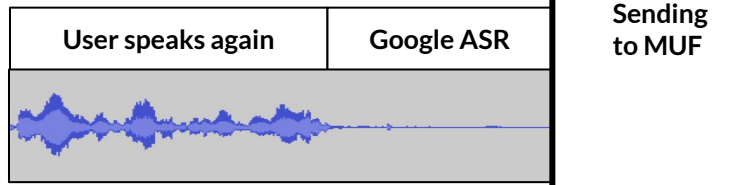
Google ASR  
output

Sending  
to MUF

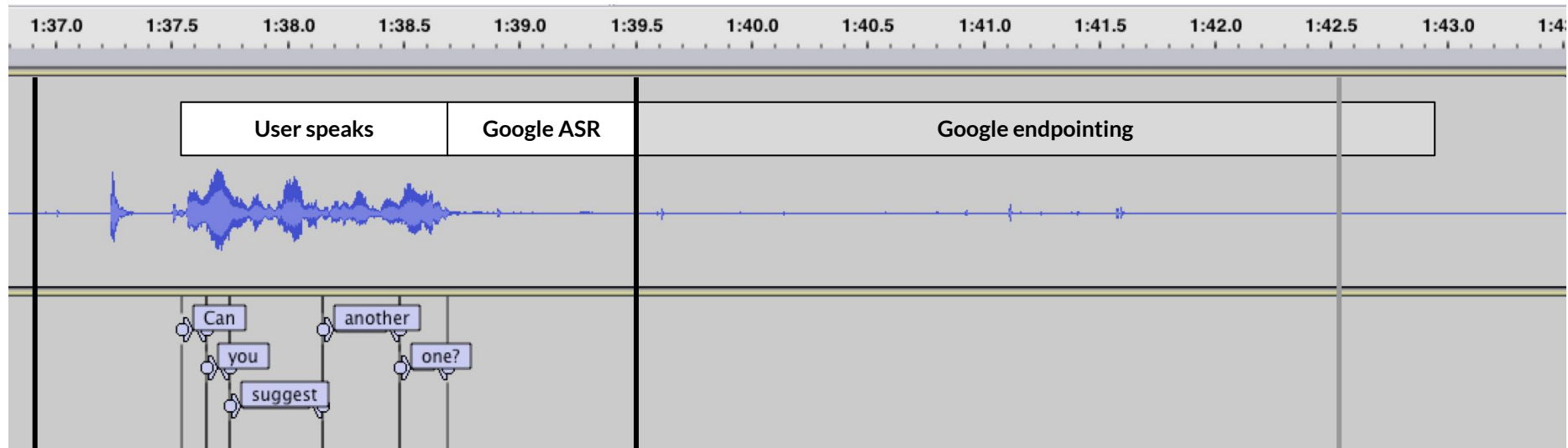


Pointer down

Google ASR output



Second Google ASR output



Pointer down

Google ASR output

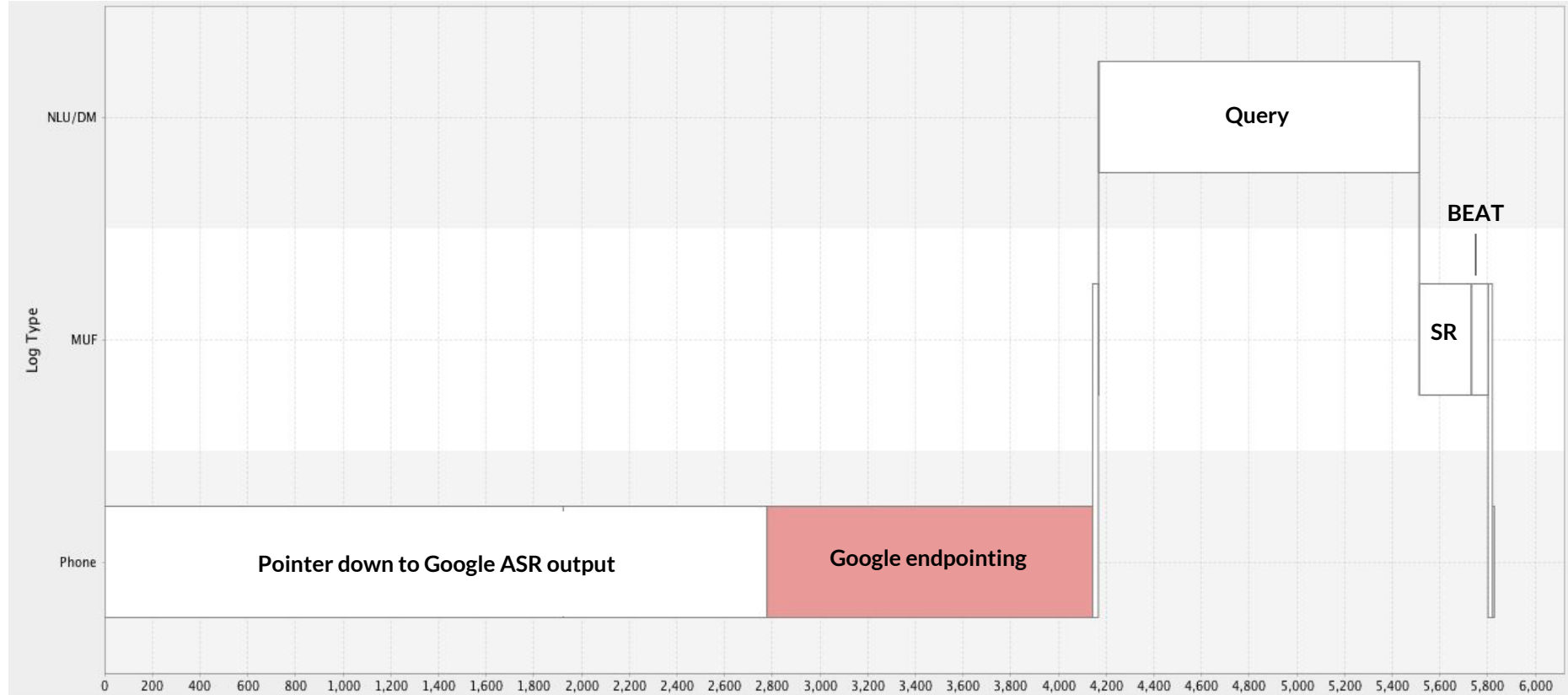
User speaks again

Google ASR

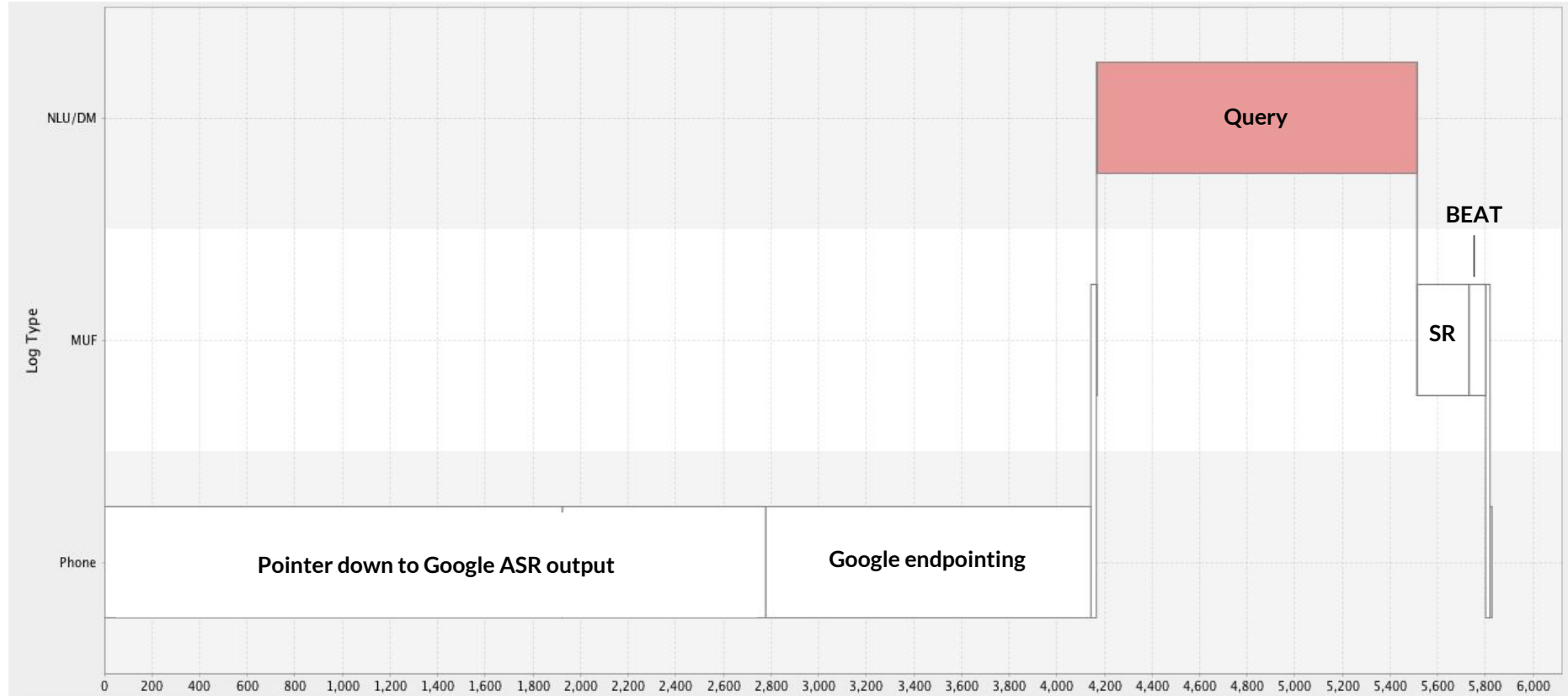
Sending to MUF

Second Google ASR output

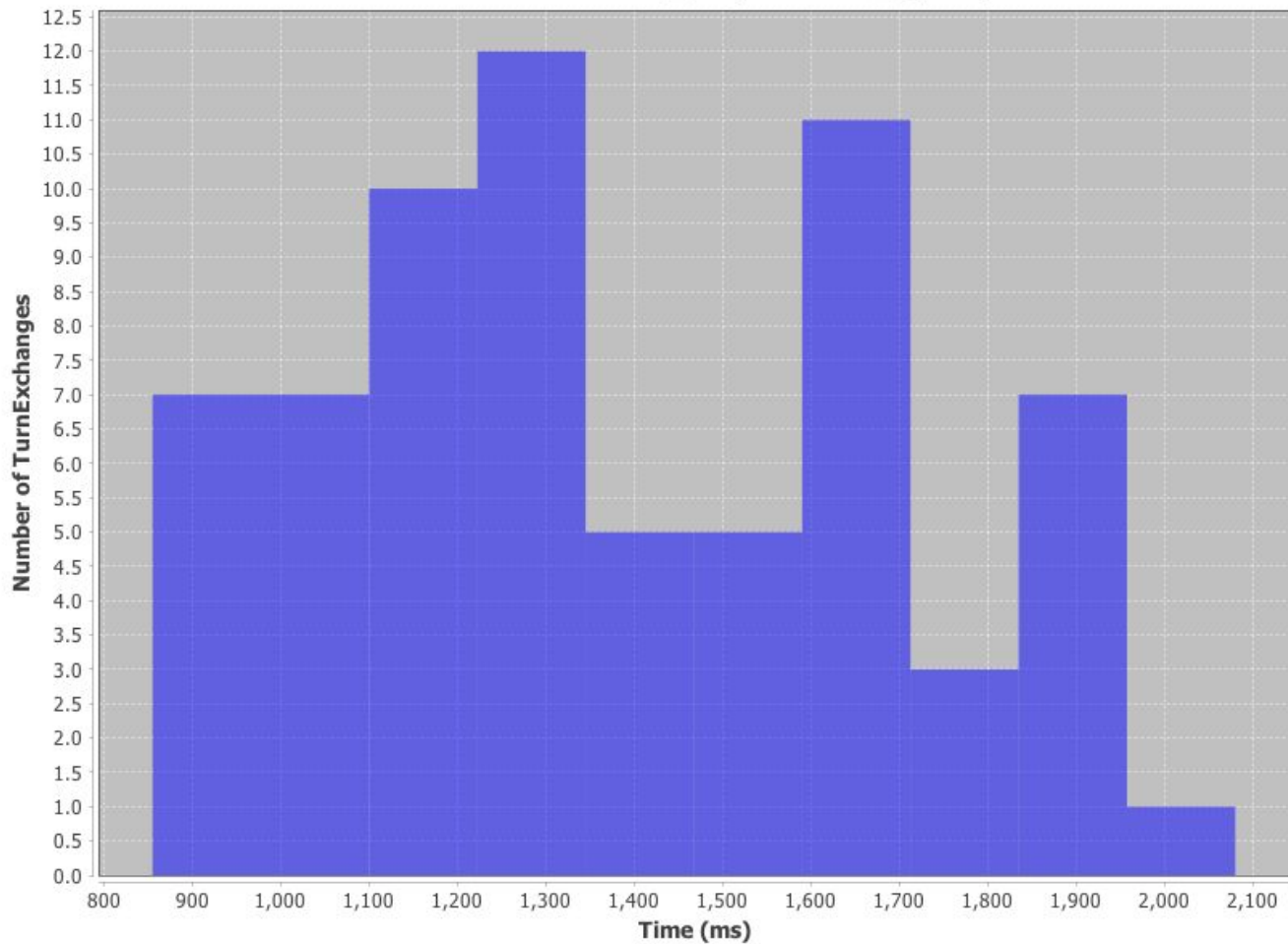
# Current Timeline (with Query)

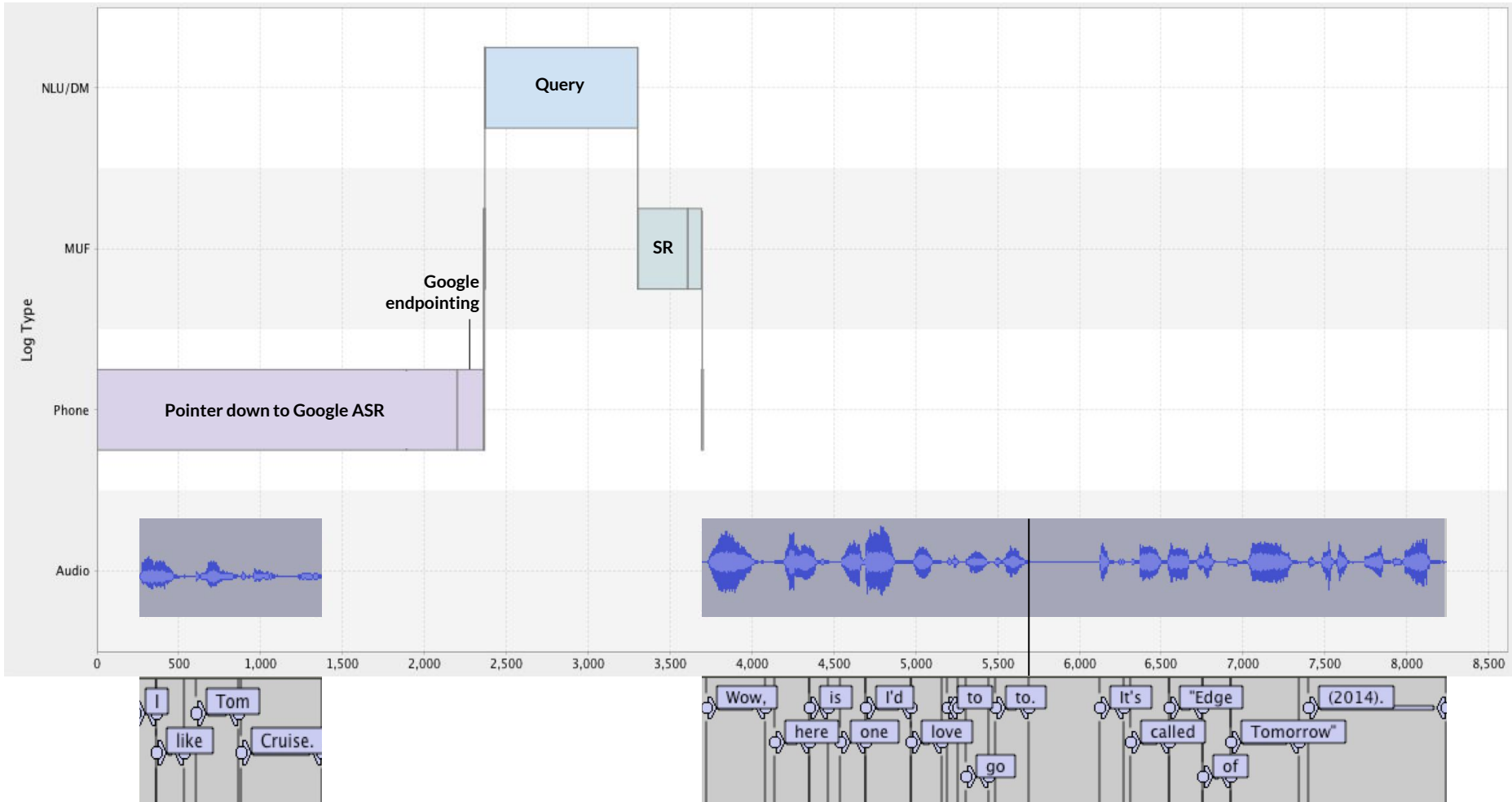


# Current Timeline (with Query)

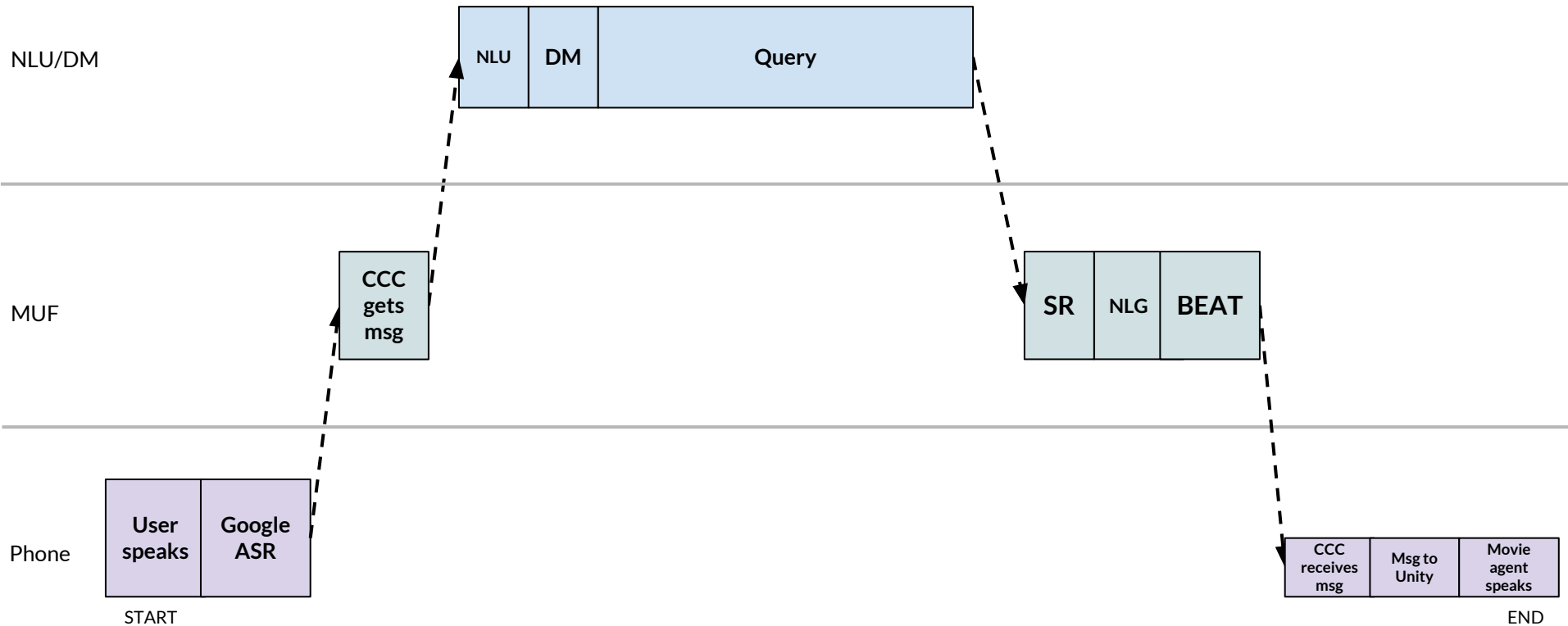


**Time Between NLUDM\_DM, NLUDM\_Query**



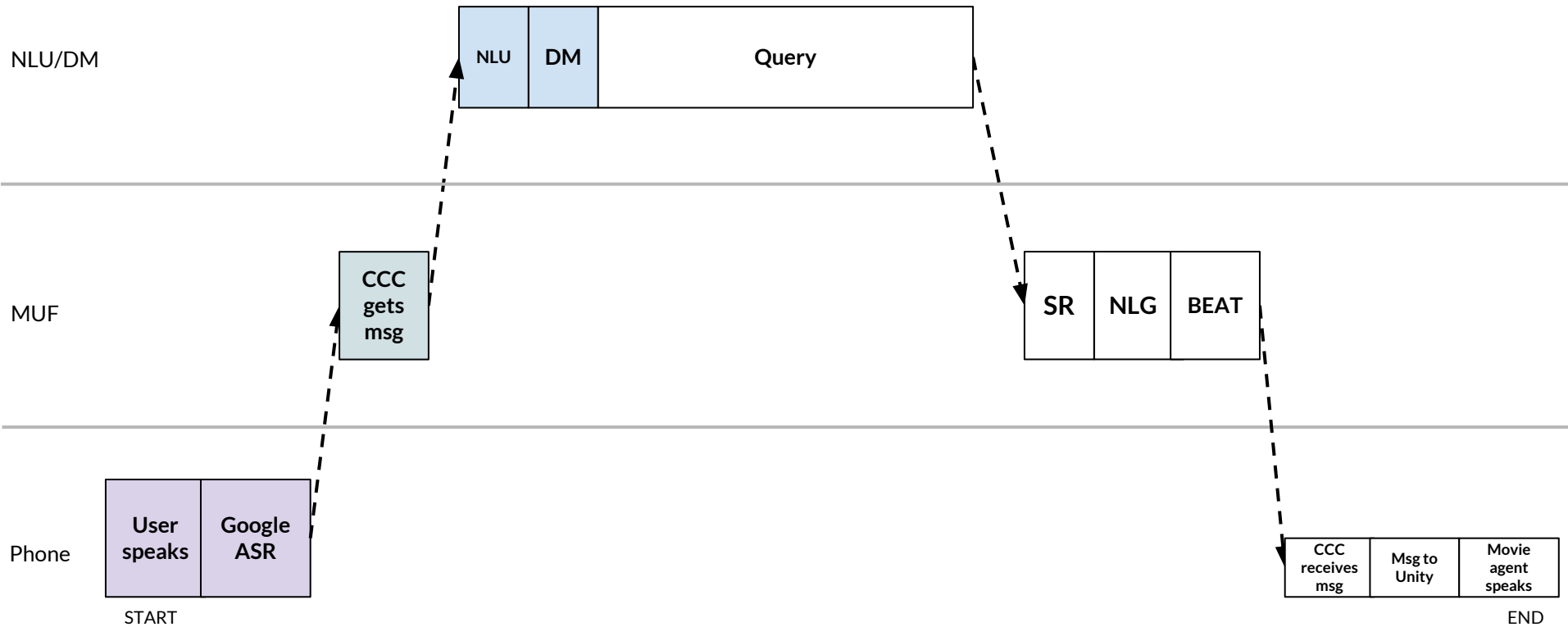


# Proposed Flow of Events

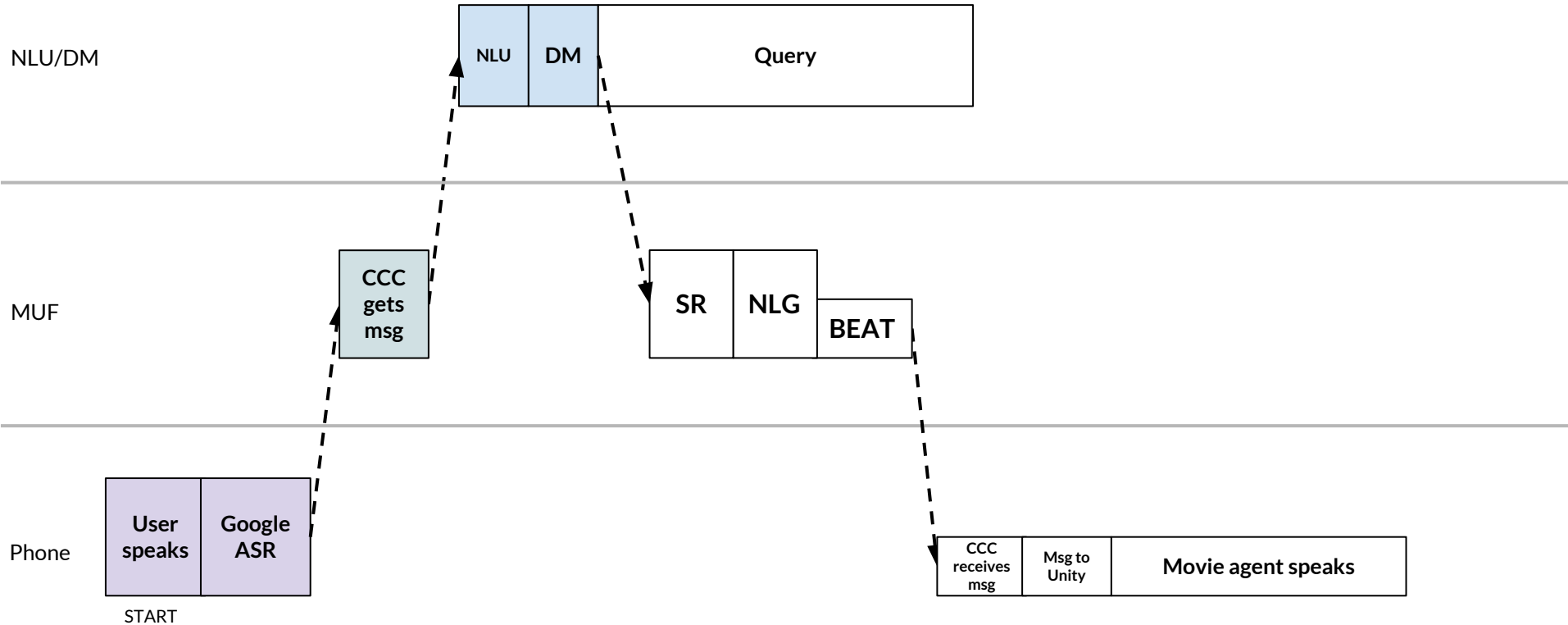




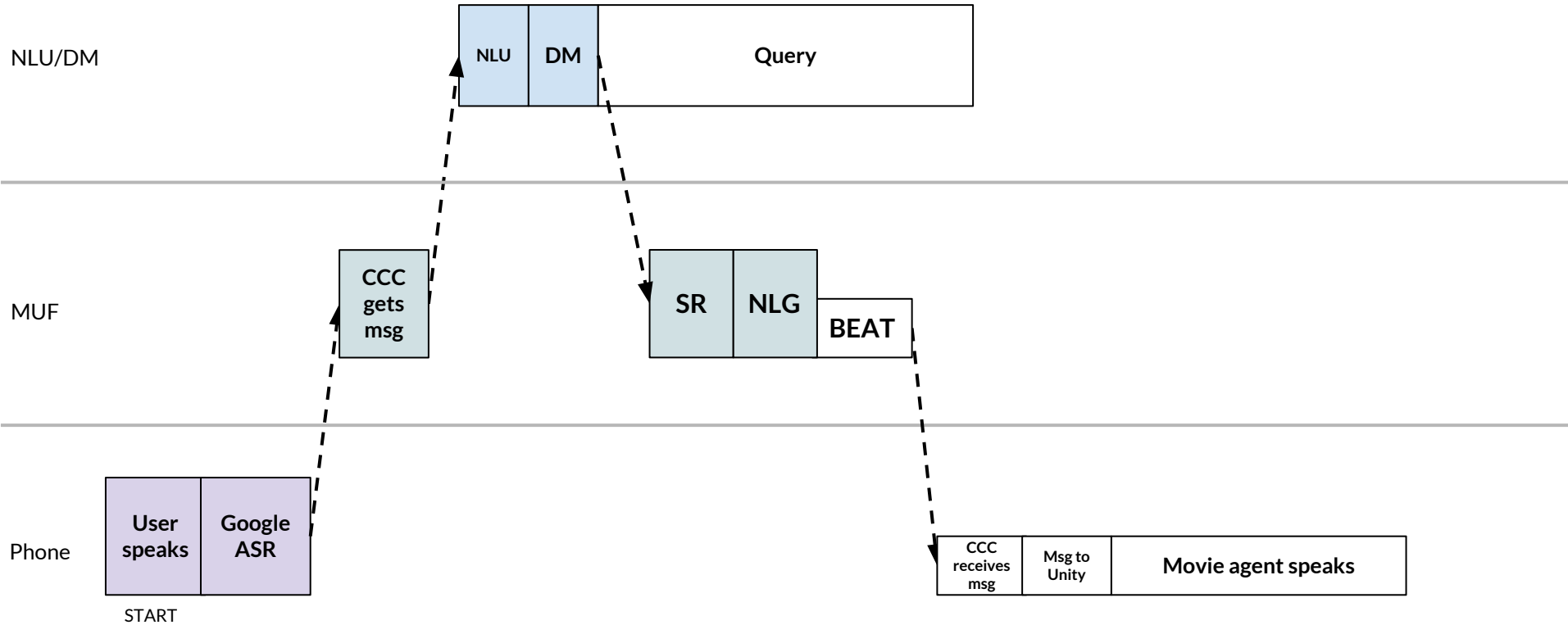
# Proposed Flow of Events



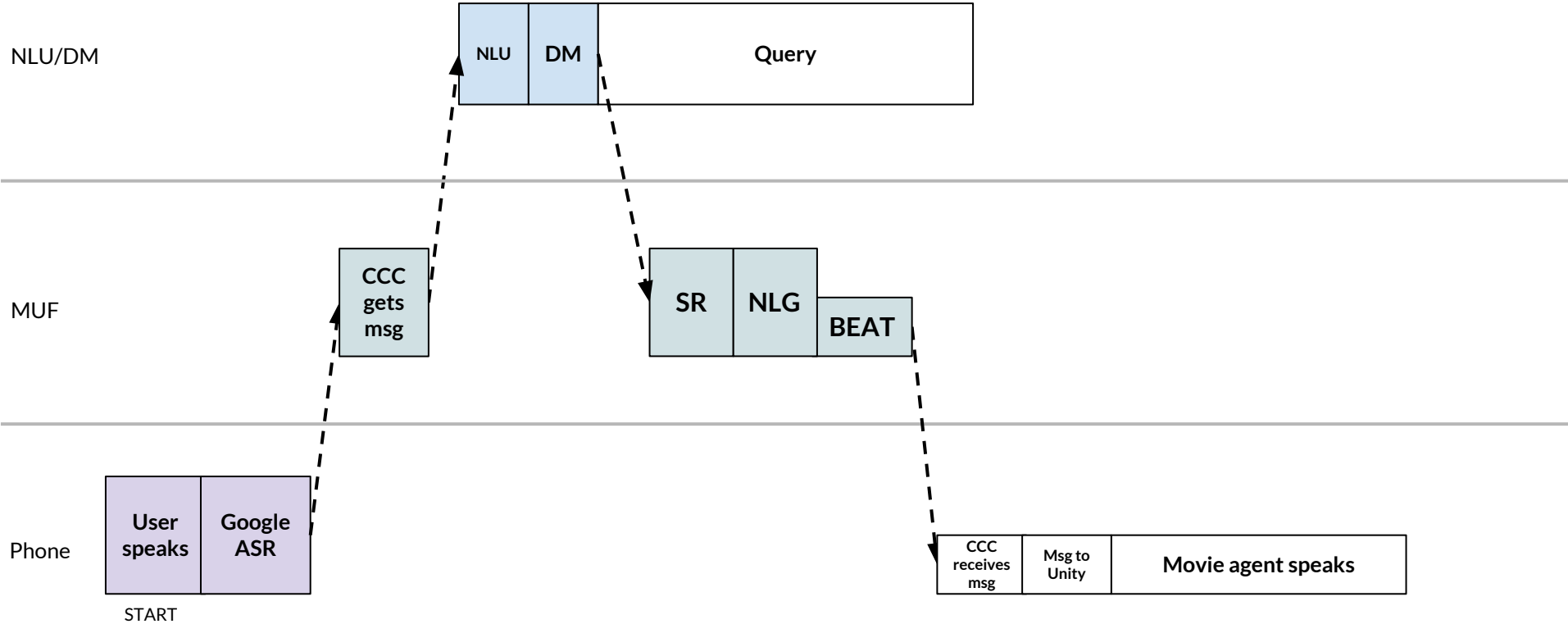
# Proposed Flow of Events



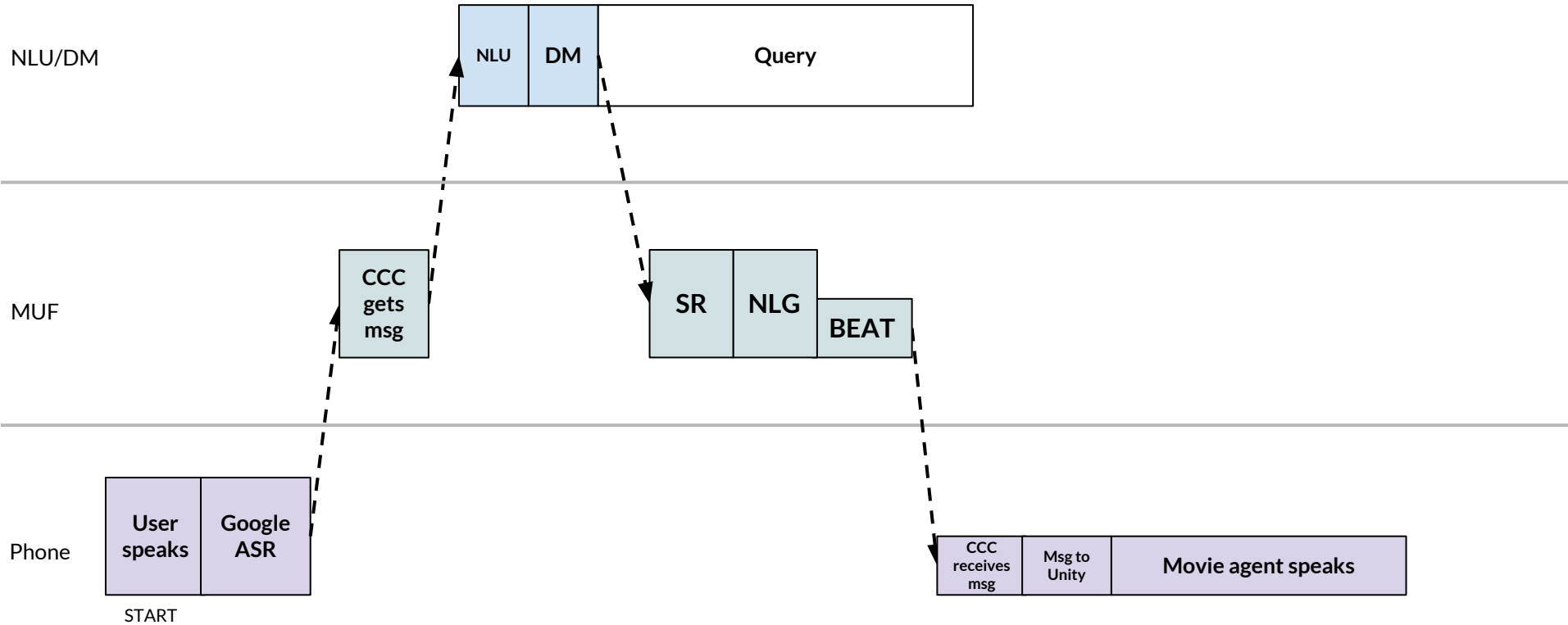
# Proposed Flow of Events



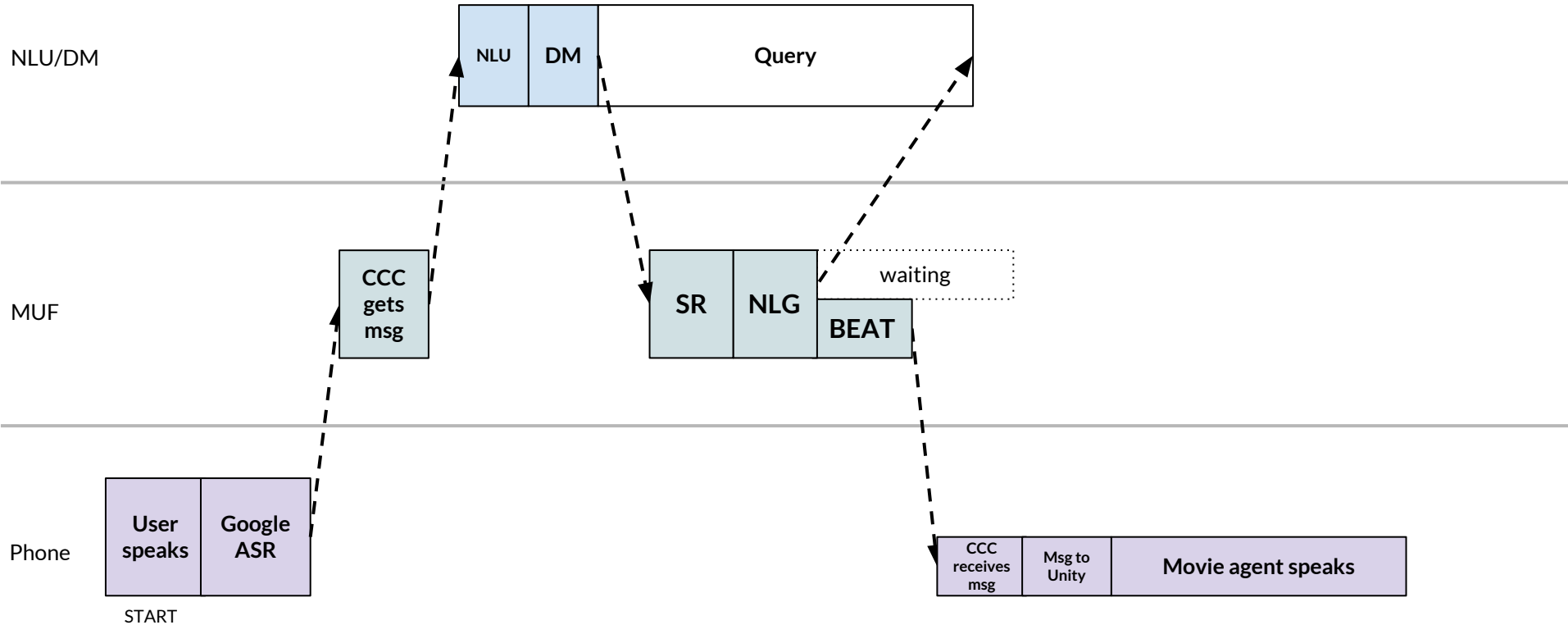
# Proposed Flow of Events



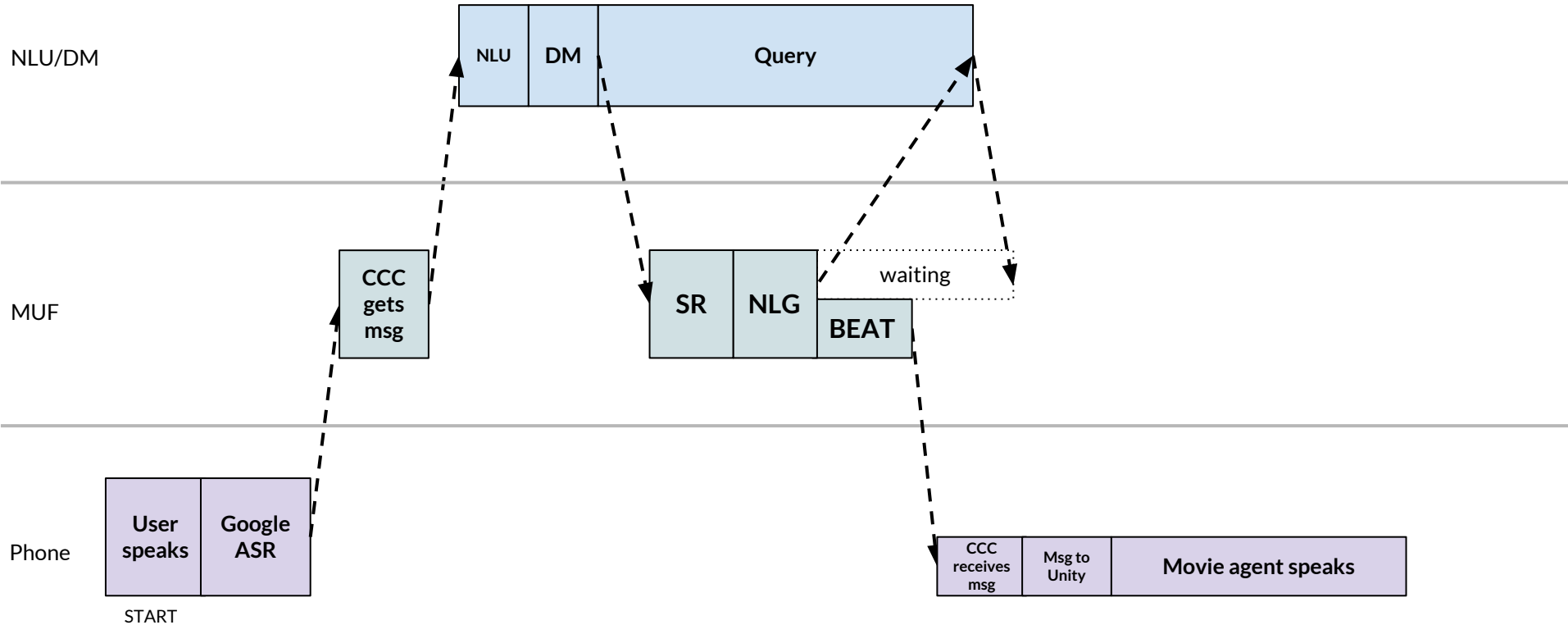
# Proposed Flow of Events



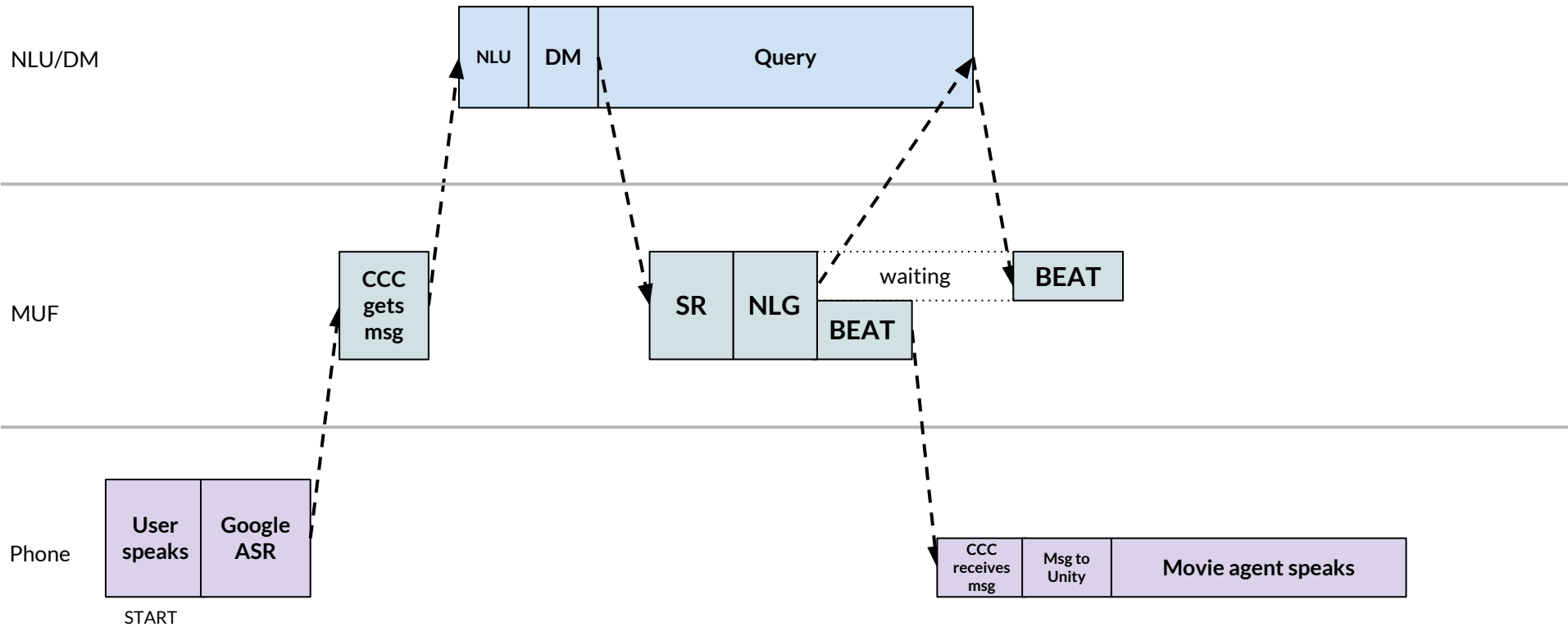
# Proposed Flow of Events



# Proposed Flow of Events

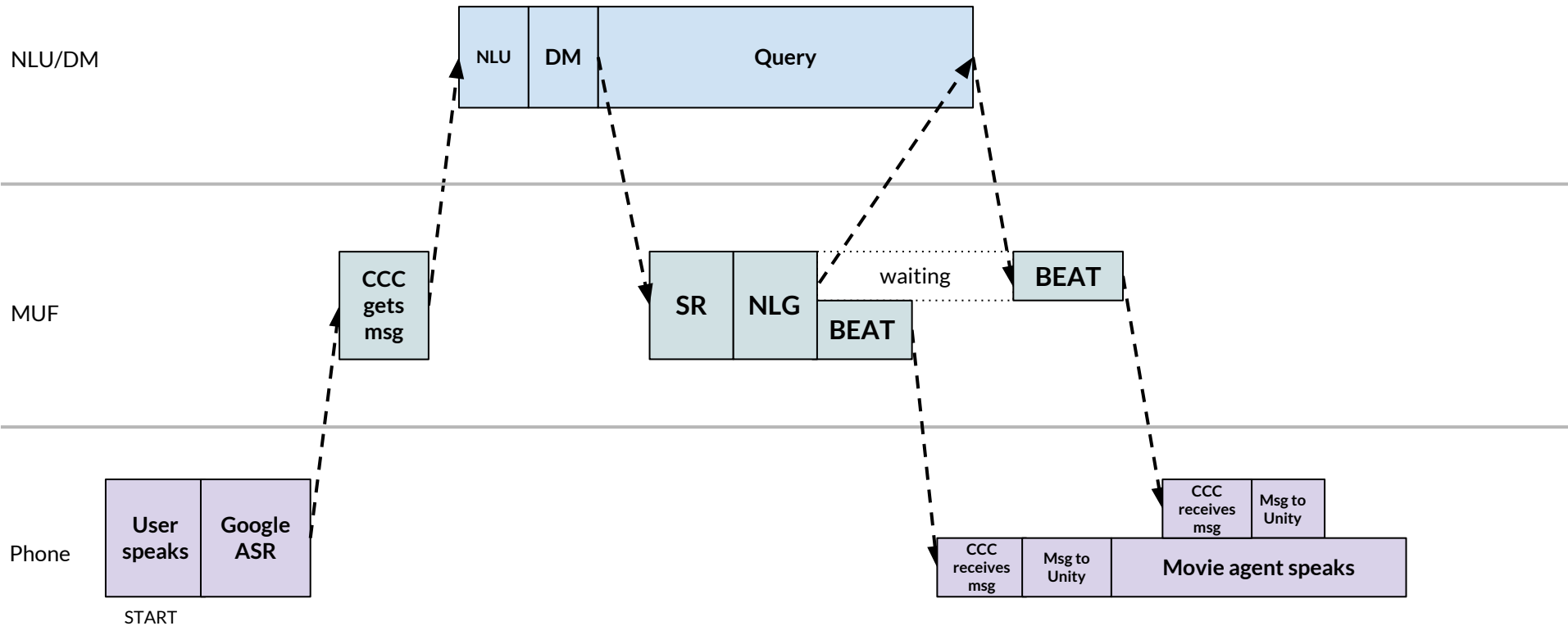


# Proposed Flow of Events

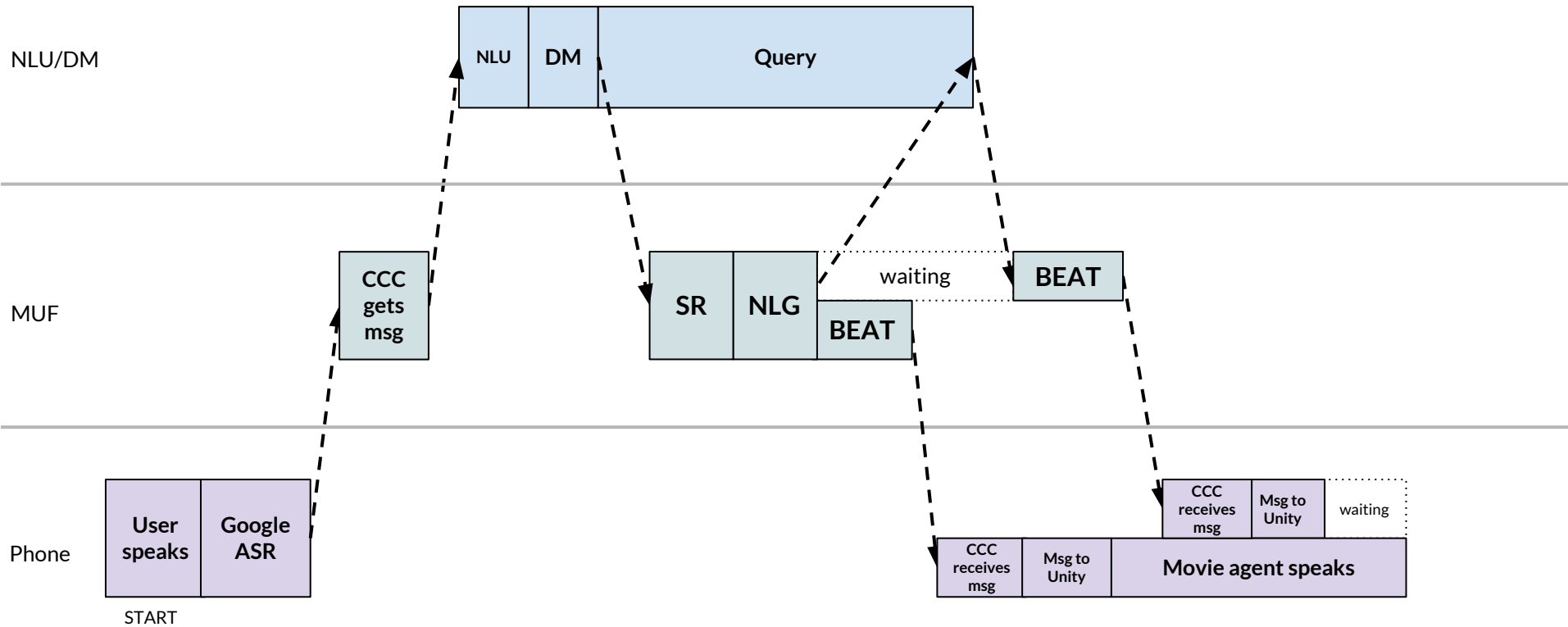




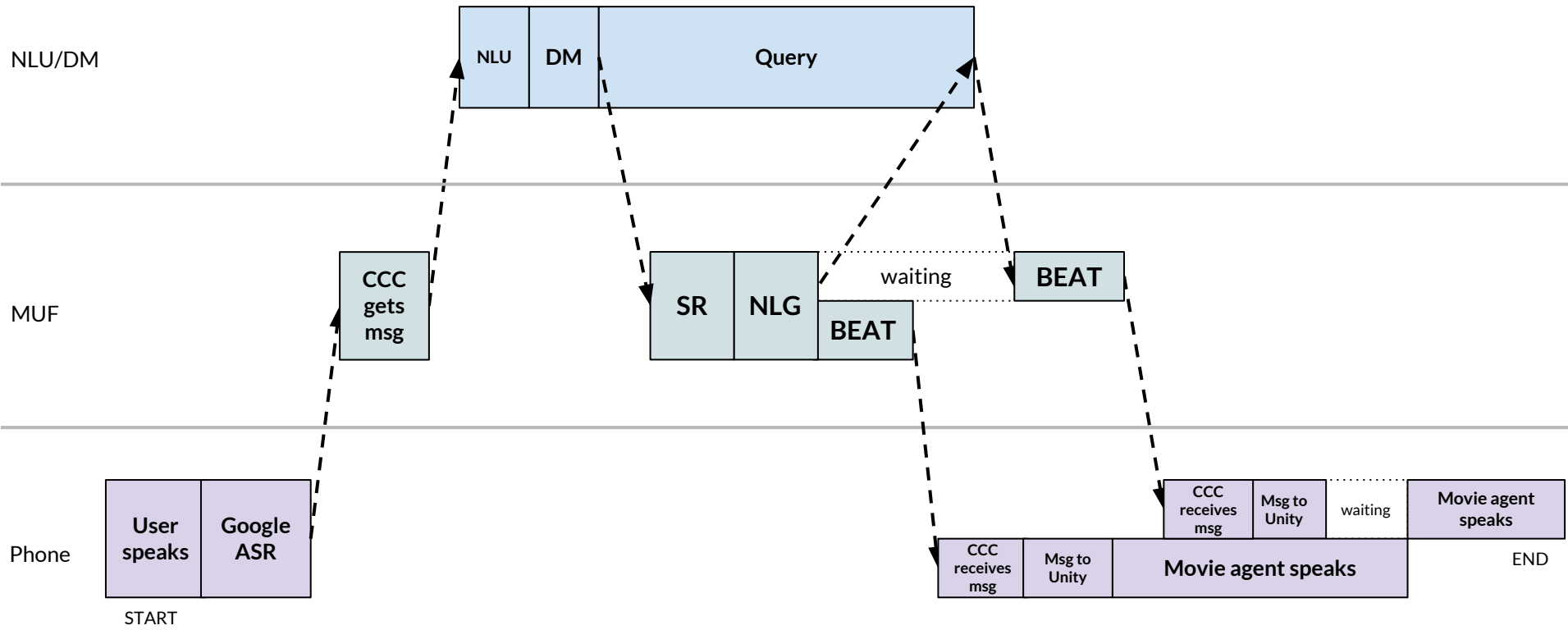
# Proposed Flow of Events

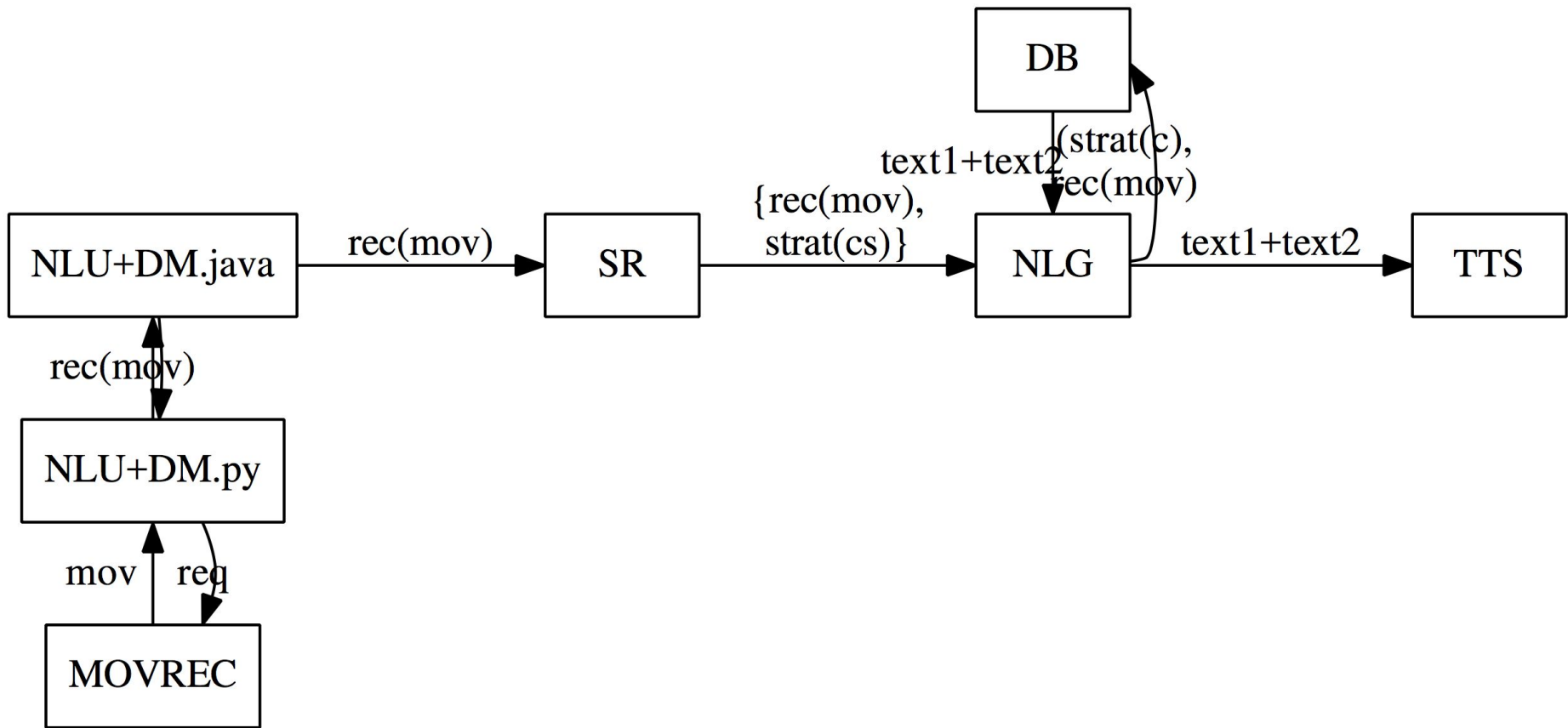


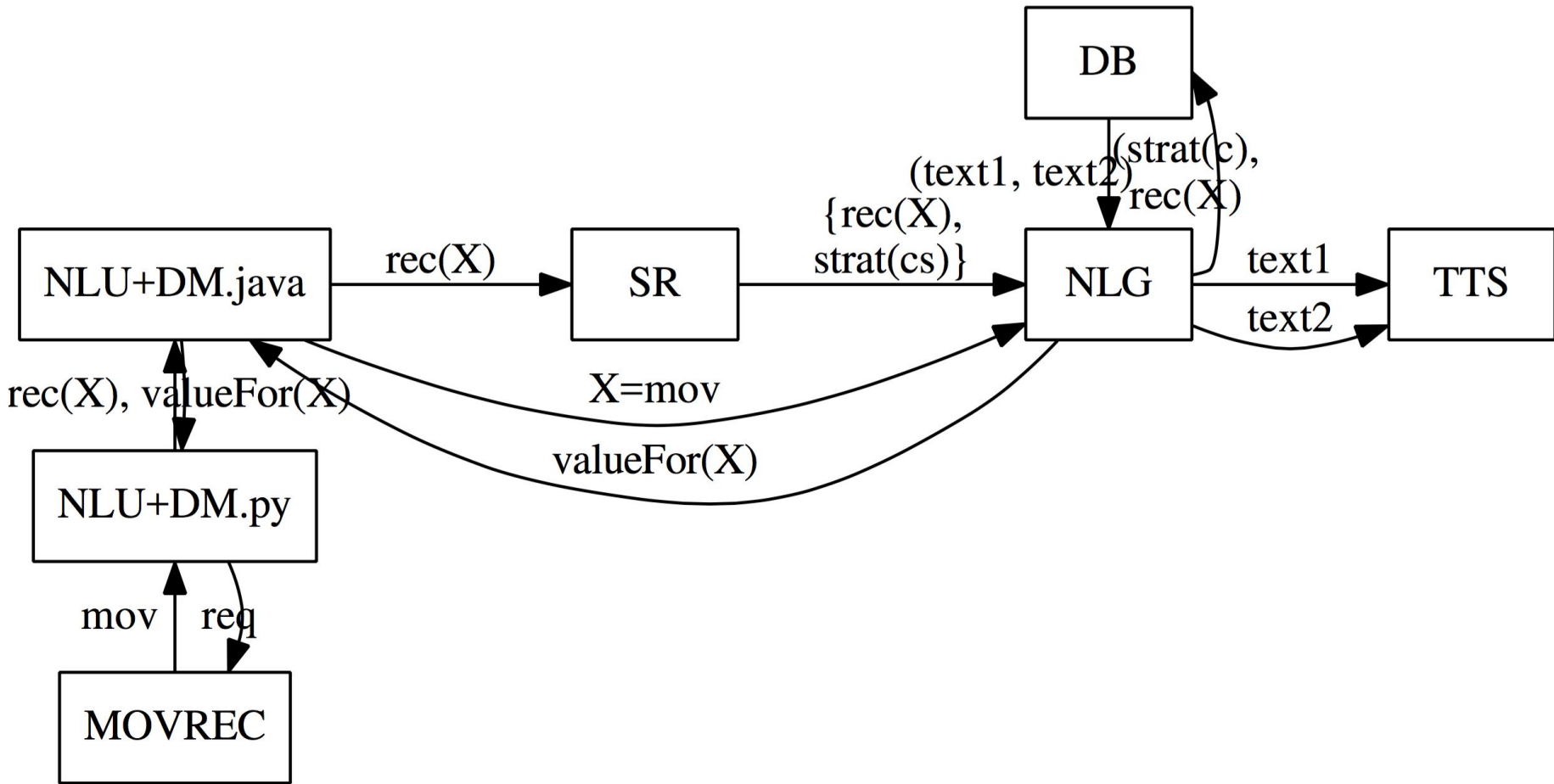
# Proposed Flow of Events



# Proposed Flow of Events







# Conclusion

- Analysis of TurnExchanges shows 2 significant delays
  1. Google endpointing
    - Solution: send first result and ignore subsequent results
  2. Query to movie database
    - Solution: output response incrementally
    - Necessary changes:
      - DM - handle 2 queries (underspecified variable, actual value)
      - NLG - split response into 2 patterns; get variable
      - TTS - enqueue recommendation sentence to avoid interruptions
- Secondary analysis to be conducted after changes are made