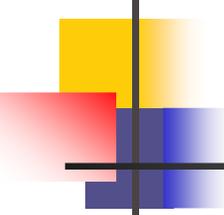


# Text to Speech Technologies for Mobile Telephony Services

---

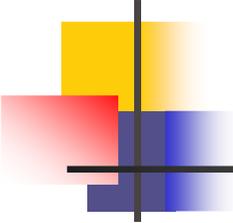
Paulseph-John Farrugia



# Presentation Outline

---

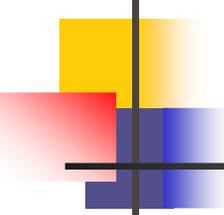
- Research context is that of part-time M.Sc. in Computer Science focusing on NLP for the Maltese Language
- A brief overview of the Text to Speech (TTS) will be given
- Focus of research presented



# Introduction

---

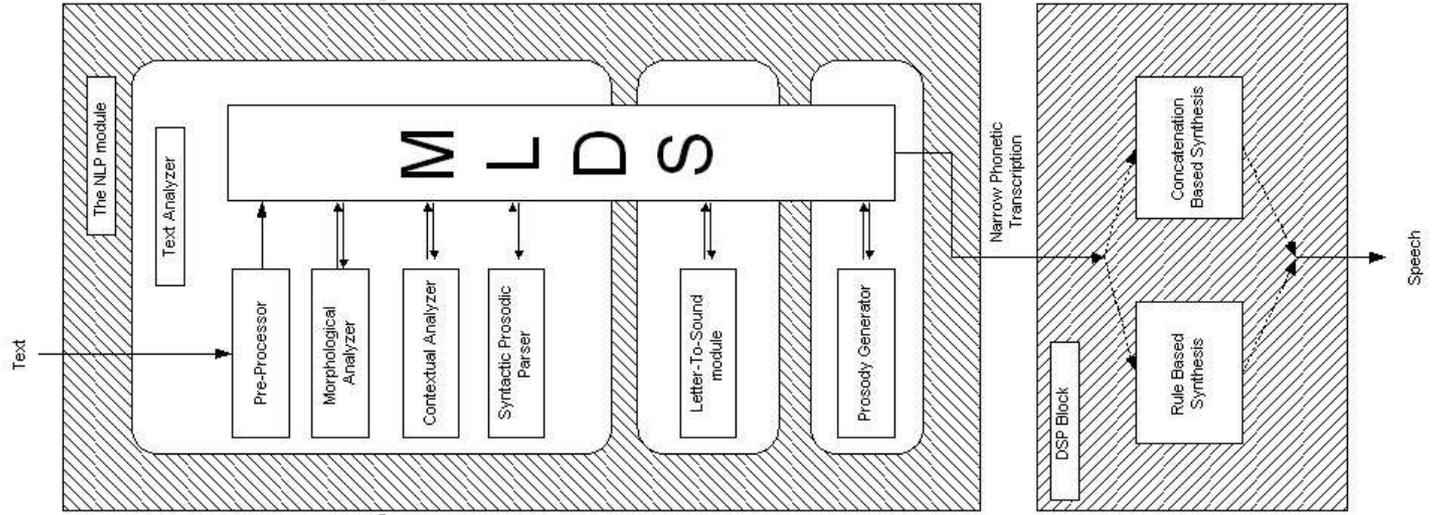
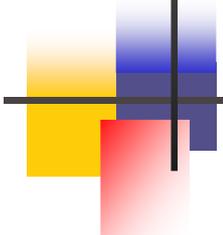
- A definition of TTS
  - The transformation from arbitrary text to audible speech
- Not a straightforward task
- Applications
  - Aid for the visually impaired
  - Language education
  - Man-Machine Interaction
- Intelligibility and Naturalness

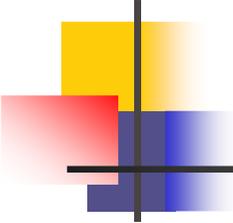


# Structuring the Task

---

- TTS requires both NLP and DSP techniques
- Task can be viewed as a transducer with 2 main blocks containing NLP or DSP modules
- The NLP block analyses the text in order to provide a “narrow phonetic transcription”
- The DSP block utilizes this transcription in order to generate the corresponding sound output

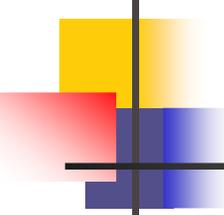




# The NLP Block

---

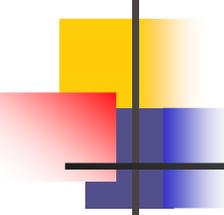
- The more language specific block
- Consists the following modules:
  - Pre-processor
  - Morphological analyser
  - Contextual analyser
  - Syntactic Prosodic Parser
  - Letter to Sound module
  - Prosody generator



# Pre-Processor

---

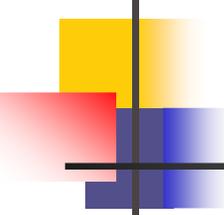
- Recognizes and spells out acronyms (consider CNN, SARS, CSAW)
- Spelling out of numbers and figures (consider 25, Lm 25.00, 15 ktieb)
- Resolve punctuation ambiguity (consider use of . for abbreviations or as a sentence marker)



# Morphological Analyzer

---

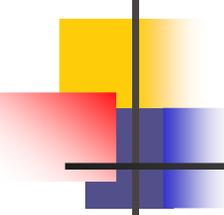
- Utilizes lexical information in order to carry out a morphological decomposition of individual words
- Proposes possible parts of speech for each word



# Contextual Analyzer

---

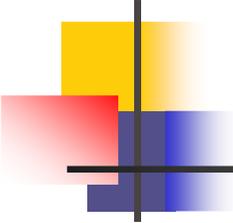
- Utilizes grammatical information and the previous results in order to derive the most likely parse for the input
- In this context, a full input parse is not a necessity for reasonable results



# Syntactic Prosodic Parser

---

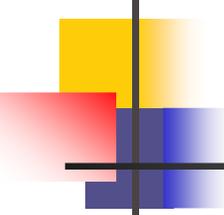
- Organizes the lexically analyzed word list into identifiable boundaries (sentences, clauses, phrases)
- Approaches
  - Grammar-driven
  - Heuristics
  - Corpus based



# Letter to Sound Module

---

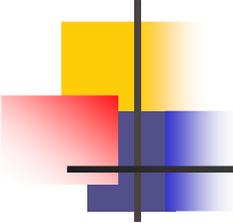
- Derives the phonetic transcription for the identified word list
- Not a simple dictionary lookup process:
  - Morphological derivations
  - Heterophonic homographs
  - New words
- Dictionary and Rule Base Approaches



# Prosody Generator

---

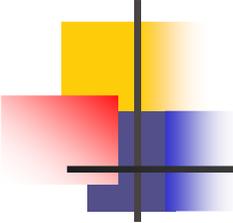
- Prepares the identified phonetic transcription and prosody contours for processing by the DSP block
- Prosody an important component for naturalness



# The DSP Block

---

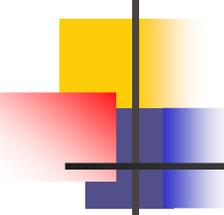
- Counterpart of the human speech reproduction system
- Responsible for transforming the narrow phonetic transcription into audible speech



# Rule-Based Synthesis

---

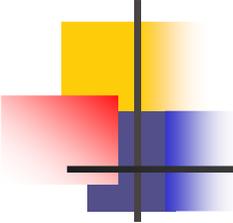
- Output generated through formant synthesis
- Speaker independent
- Low quality due to task difficulty



# Concatenation-Based Synthesis

---

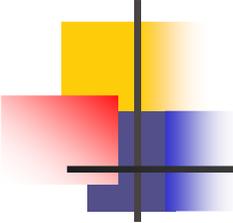
- Output generated through the concatenation of speech units (typically diphones)
- Speaker dependent
- Higher quality



# Project Focus

---

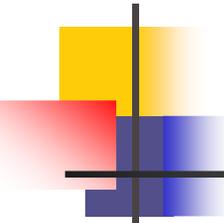
- Project deliverable is a TTS system with:
  - SMS text messaging as an input domain
  - Maltese language support
  - NLP Framework for enhanced prosody processing
- Focus is on the NLP side of TTS, in particular the prosody aspect which is very important for naturalness and is still an open research area even in linguistic circles



# Project Status

---

- Literature review
- Consideration of possible approaches for prosody analysis



# Conclusion

---

- Question time