For use by the Project lecturer	Approved		Revision required	X

Feedback

Thanks for the improvements, but you have also now created some confusion by removing some requirements (changing their details).
1. It appears you are saying "speech synthesis will be done by student (Req 3), but text to speech and linguistics processing by a library (Req 1)". The first principles design contribution then becomes vague.
Also, the core own contribution goes into requirement 1.
2. Note, use of library functions is not allowed
3. I'm missing a specification for how quickly the processing needs to be done (with appropriate motivation)

First revision.

Included initials and surname.

Revision required
(See Project Clickup site for submission deadline)

To be completed by the student

PROJECT PROPOSAL 2022

Project no	AO5	Revision no	1

Title	Surname	Initials	Student no	Study leader (title, initials, surname)
Mr	**Ramedies**	**I**	16023405	Mr A Oloo

Project title

A school intercom system that filters inappropriate phrases and plays emails as announcements in the voice of the sender.

Language editor name	Language editor signature
N. Ramedies	
Student declaration I understand what plagiarism is and that I have to complete my project on my own.	**Study leader declaration** This is a clear and unambiguous description of what is required in this project. Approved for submission (Yes/No)
Student signature	Study leader signature and date

Section was language edited.

1. Project description

What is your project about? What does your system have to do? What is the problem to be solved?

You can remove this motivation now - we are beyond this point.

The problem addressed in this project, is the lack of physical contact between teachers and students in schools due to the recent change in schooling systems. Teachers that make announcements and students that hear them, are often not able to be in the same physical location. Employing an individual to collate announcements would not be an efficient solution to this problem. Thus the problem addressed in this project, is to develop a school intercom system that will make announcements on behalf of the teacher, eliminating the need for announcers or any human intervention.

The concept is to have a system that receives a text-based email from a teacher, then announces this email to the students in the voice of the teacher that sent the email. This may be accomplished by integrating natural language processing techniques, speech synthesis, and digital signal processing, into a single intercom system. The system should also take account of any inappropriate phrases received from the email and should be able to play the announcement over an integrated speaker, so that students in a classroom are able to hear it.

Added a layer of challenges that reflect the functional analysis of the system.

2. Technical challenges in this project

Describe the technical challenges that are *beyond* those encountered up to the end of third year and in other final year modules.

2.1 Primary *design* challenges

One of the design challenges would be, accounting for multiple users' voices. Another is, hardware allocation of the system resources. The overall system should be designed to work in real time, so all training of artificial intelligence systems, if used, will have to be conducted prior to final product compilation. The linguistic analysis technique used will have an affect on the processing time of the overall implementation, and each technique has benefits over others. Selection of a specific technique, or development thereof, will be seen as a design challenge.

2.2 Primary *implementation* challenges

A system implementation challenge is, managing waveforms and assigning it to analyzed text correctly. Testing results of synthesized audio, when compared to natural language audio, might be difficult to achieve, unless extra functional components are developed. It also might be challenging to develop bandpass filters with very narrow passbands, depending on the support that the development platform provides. Another implementation challenge would be within the natural language processing and linguistics analysis algorithms. The English language contains many rules and it might prove challenging to combine multiple rules without making design errors within the algorithms.

Section was changed to add an extra layer that includes functions that is contained within the algorithms.

3. Functional analysis

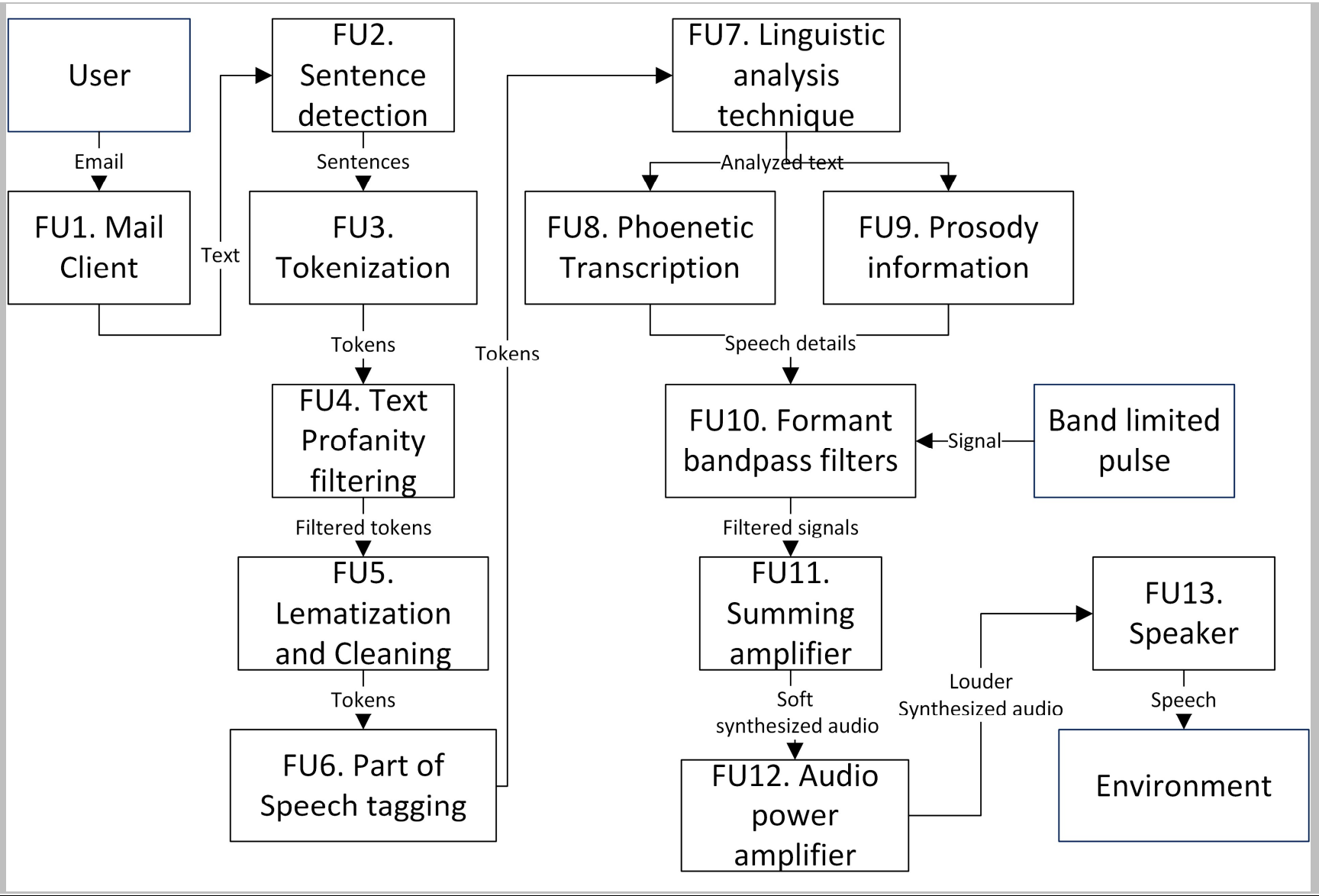
3.1 Functional description

Describe the design in terms of system functions as shown on the functional block diagram in section 3.2. This description should be in *narrative format*.

The user input will be an email text announcement that is sent to the system via the internet. FU1 is a mail client that will receive these emails. The text is extracted from the email, whereafter, natural language processing (NLP) and linguistics analysis is performed on the text from FU2 to FU7. The process involves detecting sentences (FU2), and creating tokens from those sentences (FU3). These tokens are run through a profanity filter (FU4) that removes all inappropriate phrases contained within a set of tokens, by checking the tokens against a database of profane words. The output of the profanity filter will be a set of tokens with some marked to be inappropriate, if applicable. The set of marked tokens will then go through a lemmatization and cleaning process, in FU5, where sentences are correctly formatted. Each token is then tagged with its part of speech (FU6), relevant to the whole sentence. A linguistics analysis technique, such as a rule-based technique, is then performed on the output, in FU7, to determine the meaning of the tagged sentences. The analyzed text is sent to FU8, a phonetic transcription function, that will append phonemes to every part of the word, representing what it will sound like. The analyzed text is simultaneously sent to a prosody generator (FU9), which determines the prosody information, i.e. the pitch, loudness, and duration of the words. The information generated from the combination of FU8 and FU9 are then used in FU10 and FU11, that make up a waveform generator or synthesizer. The synthesizer artificially creates a sound signal from the retrieved prosody information using formant synthesis. Formant synthesis combines 5 different frequencies (formants), of a user's voice, utilizing digital bandpass filters (FU10), into a single voice signal, by using a summing amplifier (FU11) . Depending on the strength of the synthesized signal, it may be amplified through an audio power amplifier (FU12). The synthesized speech would then be output as an announcement through FU13, a speaker. The output should be heard in the classroom environment by students.

Block diagram reflects what the linguistics analysis and NLP will consist of. As well as the speech synthesis approach functions.

3.2 Functional block diagram



4. System requirements and specifications

These are the core requirements of the system or product (the mission-critical requirements) in table format IN ORDER OF IMPORTANCE. Requirement 1 is the most fundamental requirement.

	Requirement 1: the fundamental functional and performance requirement of your project	**Requirement 2**	**Requirement 3**
1. Core mission requirements of the system or product. Focus on requirements that are core to solving the engineering problem. These will reflect the solution to the problem.	The system should be able to convert an arbitrary text input to speech information in natural language.	The text input will have to be processed so that it is correctly perceived by the students, i.e. it should be checked for correctness in punctuation and spelling.	The announcement has to be broadcasted in a synthetic voice that sounds identical to the sender's voice, in terms of the pitch, loudness, length, and strength.
2. What is the target specification (in *measurable* terms) to be met in order to achieve this requirement?	A linguistics analysis module produces word error rates (WER) of 30% or below after processing text. Pitch and gain of a synthesizer should be 100% accurately aligned with the output speech information of the linguistics analyzer.	The overall linguistic performance (OLP) of the spell checker should be approximately 90%, and the punctuation checker should account for all the beginnings and endings of sentences.	The fundamental frequencies, amplitude, duration and amplitude dynamics of the formants in a speaker's voice should be band limited. All aspects should not differentiate more than 2% of the original voice.
3. Motivation: *how* or *why* will meeting the specification given in point 2 above *solve the problem*? (Motivate the *specific* target specification selected)	Developed speech-to-text transcribers of human voices yield WERs between 15 to 40%. It is expected that synthesized speech, with correct prosody, will lie within this range, utilising these transcription tools.	A 90% metric aligns with already developed spell checkers, in terms of recall, precision, and suggestion of words. The reason for this is that there are a large percentage of homonyms, and often spell checkers cannot pick up on this.	Adults can differentiate a 1% change in 4000Hz tones, which equates to a 4Hz difference. Males and females have much lower fundamental frequencies at a maximum of 155Hz and 255Hz respectively. A 2% change in this metric corresponds to 3-5Hz.
4. How will you demonstrate at the examination that this requirement (point 1 above) and specification (point 2 above) has been met?	Text data will be fed into the intercom system, wherein a linguistics analysis function will provide a synthesizer with speech information to generate. A separate speech to text transcriber will display the perceived announcement and WER.	A data set of sentences will be input to a spell and punctuation checker. The output of the data set sentences should be displayed along with the results of OLP.	Recordings of multiple announcers' voices will be played, in conjunction with the synthesized speech for auditory comparison. Waveforms of the speech will also be displayed and fundamental frequency ranges will be compared.
5. Your own design contribution: what are the aspects that *you will design and implement yourself* to meet the requirement in point 2? If none, *remove this requirement.*	A linguistics analysis module, will be implemented to generate speech information. A simulated transcriber platform to retrieve results, will be designed and implemented.	A software implementation of a spell checker and a basic punctuation checker, will be developed and implemented.	A formant speech synthesizer, consisting of digital filters, will be designed and implemented. A prosody generation function to serve as input to the speech synthesizer, will be implemented.
6. What are the aspects to be taken off the shelf to meet this requirement? If none, indicate "none"	A microcontroller board with adequate processing power. A speech recognition library to perform speech to text transcription for results. An audio recording device will be used to store system output.	A dictionary of words to replace incorrectly spelled words.	A microcontroller board, capable of digital signal processing, will be purchased. Software to track formant frequencies of the announcer will be used to assist with design.

Specified the exact parameters that will be considered for every synthesized signal.

More accurate description to clear up any thing that might not make sense.

Better technical term than accuracy.

Specifications reflect the detailed mission requirement.

Includes reasoning of 90% choice.

Motivates the new specifications.

Adding motivation for the 30% WER choice.

This feature is not simulated.

Intention not clear.
You are saying "speech synthesis will be done by student, but text to speech and linguistics processing by a library". What will then be the first principles design contribution?

No libraries allowed.

No, not allowed.

A separate system will be implemented that will show the results of synthesized speech.

Clarified the design contribution.

Added required tools.

More accurate description.

Includes the type of synthesizer considered.

A software will be used to obtain design parameters for the filters from audio recordings. Then the synthesizer uses those parameters to develop a voice.

System requirements and specifications page 2

This will indeed be compared against a personal database.

	Requirement 4	Requirement 5	Requirement 6
1. Core mission requirements of the system or product. Focus on requirements that are core to solving the engineering problem. These will reflect the solution to the problem.	The message that is announced, should be filtered of all inappropriate phrases. The system should consist of a comparative database and should manage to handle derivations of profanities.	A mail system should receive emails from a user.	
2. What is the target specification (in *measurable* terms) to be met in order to achieve this requirement?	The accuracy, recall and precision of a large sample set, should be above 90% each, to ensure that the profanity filter is accurate enough.	The text data that is sent by the user in the email, should be correctly reflected by the receiving end of the system, with no changes made to the input.	
3. Motivation: *how* or *why* will meeting the specification given in point 2 above *solve the problem*? (Motivate the *specific* target specification selected)	~~A 100% profanity filtering margin is unachievable.~~ A 90% margin ensures that blatant profanity is removed. This relies on the assumption that teachers are responsible in what they announce.	To ensure that the mail system receives emails correctly, it must prohibit the receiving end from making any changes to the user's input.	
4. How will you demonstrate at the examination that this requirement (point 1 above) and specification (point 2 above) has been met?	A large data set of phrases, containing profanity and False Positives, will be input into a simulated profanity filter. The outcome of accuracy, recall and precision will be measured by the models' filtered output, compared to the expected output.	A display screen should indicate notification upon receiving the email. The output of the text should be displayed and confirmed by the user.	
5. Your own design contribution: what are the aspects that *you will design and implement yourself* to meet the requirement in point 2? If none, *remove this requirement.*	A text-based profanity filter will be developed and implemented.	Implementation of an email text display feature.	
6. What are the aspects to be taken off the shelf to meet this requirement? If none, indicate "none"	None.	A Wi-Fi module, capable of Internet connection, and an LCD screen, for debugging and display purposes, will be purchased. An email library will be used for the implementation of the mail client.	

Changed from 94% to align with research papers.

Not a relevant remark. Is the TARGET not to remove ALL profanity?

Reason for 90% selection explained.

System requirements and specifications page 3

	Requirement 7	Requirement 8	Requirement 9
1. Core mission requirements of the system or product. Focus on requirements that are core to solving the engineering problem. These will reflect the solution to the problem.			
2. What is the target specification (in *measurable* terms) to be met in order to achieve this requirement?			
3. Motivation: *how* or *why* will meeting the specification given in point 2 above *solve the problem*? (Motivate the *specific* target specification selected)			
4. How will you demonstrate at the examination that this requirement (point 1 above) and specification (point 2 above) has been met?			
5. Your own design contribution: what are the aspects that *you will design and implement yourself* to meet the requirement in point 2? If none, *remove this requirement.*			
6. What are the aspects to be taken off the shelf to meet this requirement? If none, indicate "none"			

System requirements and specifications page 4

	Requirement 10	Requirement 11	Requirement 12
1. Core mission requirements of the system or product. Focus on requirements that are core to solving the engineering problem. These will reflect the solution to the problem.			
2. What is the target specification (in *measurable* terms) to be met in order to achieve this requirement?			
3. Motivation: *how* or *why* will meeting the specification given in point 2 above *solve the problem*? (Motivate the *specific* target specification selected)			
4. How will you demonstrate at the examination that this requirement (point 1 above) and specification (point 2 above) has been met?			
5. Your own design contribution: what are the aspects that *you will design and implement yourself* to meet the requirement in point 2? If none, *remove this requirement.*			
6. What are the aspects to be taken off the shelf to meet this requirement? If none, indicate "none"			

5. Field conditions

These are the REAL WORLD CONDITIONS under which your project has to work and has to be demonstrated.

	Field condition 1	Field condition 2	Field condition 3
Field condition requirement. In which field conditions does the system have to operate? Indicate the one, two or three most important field conditions.	The intercom system has to be audible enough for a noisy indoor room environment. It should be operable from any location with internet access.	Changed field condition to align with classroom conditions.	
Field condition specification. What is the specification (in measurable terms) for this field condition?	A standard noisy environment for a classroom of students ranges between 65-70 dB.		

6. Student tasks

6.1 Design and implementation tasks

List your primary design and implementation tasks in bullet list format (5-10 bullets). These are *not* product requirements, but *your* tasks.

- The intercom system must be designed and implemented.
- An appropriate linguistics analysis technique has to be selected.
- Audio data from multiple people will have to be collected.
- Large sentence text data sets, with profanity, have to be collected and sorted.
- A physical cover for the system will have to be designed and built.
- The system should be integrated onto a PCB board.
- Simulated versions of the functions should be developed for debugging and demonstration purposes.

6.2 New knowledge to be acquired

Describe what the theoretical foundation to the project is, and which new knowledge you will acquire (*beyond* that covered in any other undergraduate modules).

- The student will be required to master theoretical background knowledge on Natural Language Processing and Speech Synthesis.
- The student will be required to acquire knowledge on Linguistics Analysis techniques.
- The student will need to learn how to work with audio data sets.
- The student will be required to learn about rulings of speech in the English language.
- The student will have to learn how various aspects of digital signals affect the auditory output of speech.

Added knowledge to reflect the more detailed FBD.