

# Innovation platform SR-FP

## Speech technology for Reporting Flemish Parliament

### Report - Summary

<b>Program</b>	
<b>Contracting authorities</b>	
<b>External consultant</b>	

## 1 Summary

This document summarizes the final report of the market consultation of ‘Speech technology for Reporting Flemish Parliament’ (SR-FP), that ran from the 22<sup>nd</sup> of September until the 27<sup>th</sup> of November for which Addestino was selected as external consultant for EWI (Department Economy, Science and Innovation of the Flemish Government). It describes the priorities and next steps of composing an RFP (Request for Proposal). EWI and the Flemish Parliament aim to improve the existing workflow of reporters and editors that transcribe meetings and speeches with an innovative solution. The objective is to make the work more user friendly without lowering the quality of the transcriptions. The conclusion of the market consultation of SR-FP consists of three parts.

As a first conclusion: the project ‘Speech technology for Reporting Flemish Parliament’ is very innovative from a user point of view. The project addresses multiple (direct or indirect) user groups and tries to relieve these groups of specific pains and solve their problems and challenges. The multitude of use-cases with high added value for the end user support this claim. In total around 50 use-cases were composed with relevant end users, of which more than half were found to be “very important”. The industry partners and knowledge institutes confirmed the innovative nature from a user point of view of these use-cases.

As a second conclusion: the project ‘Speech technology for Reporting Flemish Parliament’ is very innovative from a technological point of view. The discussions with industry experts and knowledge institutes clarified that a lot of technological support for the automatic transcription of speech to text is already possible today, and even available on the commercial market. Furthermore, over 20 use-cases were identified that carry innovation potential (and risk). Those commercially available facets mostly comprise functionalities that are already available today but need tailoring to the circumstances at the Flemish Parliament. The foremost technological challenges for ‘Speech technology for Reporting Flemish Parliament’ are automatic text-to-speech that is speaker independent and shows an accuracy measure, support of speech-to-text under parliamentary circumstances including regional accents, jargon, interruptions and discussions and a user friendly workflow for automatic recognition of speakers and introduction of new speakers.

In these endeavors, quality always has priority. As a result, fully automatic reporting seems unattainable. The reporters and editors are not to be replaced, but rather supported in the process of transcription.

As a third conclusion: because of the high amount of added value of automatic transcription of speech-to-text for a lot of user groups and the technological risk that is attached to it, a precommercial development phase is needed. During this phase, a number of precommercial procurements are launched with the goals of improving and evaluating the current performance and quality that are built around the central application.

The proposed next steps of this project include forming a detailed timeline and budget for the precommercial phase and compiling a balanced base of technology providers. However, those are not a part of the context of this document. Finally, Addestino wishes to thank EWI for the confidence that was given to us and the Flemish Parliament for the delivered expertise and effort and all partners for their competence and enthusiastic co-operation.

## 2 Use-cases

Use-cases were developed from the point of view of the future users. In a first phase, the importance of each use-case was estimated by giving scores with a technique called planning poker. This score is used as the innovation potential from the user point of view, or value. In a second phase, the risk of implementing the use-case is assessed by the market, also with planning poker. This score is used as the innovation potential from a technological point of view. The synthesis of the results is given in the next chapter.

As a(n)...	I can...	so that...	ID
Editor	open a Word-document with the transcription of one of my meetings in it	I do not have to listen to the audio and quickly compose a finished report	1.1a
Editor	open a Word-document with the transcription of one of my meetings in it	I do not have to listen to the audio and quickly compose a finished report	1.1b
Reporter	open a Word-document with the transcription of one of my meetings in it	I do not have to listen to the audio and quickly compose a finished report	1.2a
Reporter	open a Word-document with the transcription of one of my meetings in it	I do not have to listen to the audio and quickly compose a finished report	1.2b
Editor	see the percentage of accuracy of words or parts of phrases	I can efficiently check which parts I need to check manually	2.1
Reporter	see the percentage of accuracy of words or parts of phrases	I can efficiently check which parts I need to check manually	2.2
Editor	choose between alternatives of words of parts of phrases	the typing workload is lower in case of transcription errors	2.3
Reporter	choose between alternatives of words of parts of phrases	the typing workload is lower in case of transcription errors	2.4
Reporter	clearly see who is speaking for each part of the delivered text with an accuracy of 100%	I can note this in my report quickly and efficiently	3.1a
Reporter	clearly see who is speaking for each part of the delivered text with an accuracy of 80%	I can note this in my report quickly and efficiently	3.2
Editor	make a distinction between the chairman speaking for himself or as chairman	I can note this in my report quickly and efficiently	3.3
Reporter	make a distinction between the chairman speaking for himself or as chairman	I can note this in my report quickly and efficiently	3.4
System administrator	automatically introduce one time speakers in the system	speaker recognition stays up to date for externs	4.1
System administrator	automatically introduce one time speakers in the system based on their first minutes of speech	speaker recognition stays up to date for externs	4.2a
System administrator	automatically introduce one time speakers in the system based on their first minutes of speech	speaker recognition stays up to date for externs	4.2b
System administrator	automatically introduce one time speakers in the system by letting them read a text as training material	speaker recognition stays up to date for externs	4.3

System administrator	easily introduce large groups of new speakers in the system	speaker recognition stays up to date after the changes in the parliament every 5 years	4.4
Reporter	introduce editorial guidelines which are taken into account by the system	I have less work correcting mistakes	5.1
Reporter	introduce editorial guidelines which are taken into account by the system	I have less work correcting mistakes	5.1
Editor	introduce editorial guidelines which are taken into account by the system	I have less work correcting mistakes	5.2
Editor	introduce editorial guidelines which are taken into account by the system	I have less work correcting mistakes	5.2
Reporter	use a lexicon with specific political and legal language including references to institutions and organizations	I have less work correcting mistakes	5.3
Reporter	use a lexicon with specific political and legal language including references to institutions and organizations	I have less work correcting mistakes	5.4
Editor or Reporter	adjust the lexicon according to the field of the subject	adjust the lexicon according to the field of the subject	5.5
Editor	link each part of the text to the corresponding part of the audio	I can quickly look up dubious transcriptions	6.1
Reporter	link each part of the text to the corresponding part of the audio	I can quickly look up dubious transcriptions	6.2
Editor	get a live transcription and make live adjustments	I have a correct transcription after the meeting	7.1
Reporter	get a live transcription and make live adjustments	I have a correct transcription after the meeting	7.2
Editor	hardly be delayed by system failure	I never miss a deadline	8.1
Reporter	hardly be delayed by system failure	I never miss a deadline	8.2
Flemish Parliament	have subtitles for the video's published on the website of the Flemish Parliament	I can reach a broader public	9
Editor	not notice influence of sounds coming from the environment, background noise and echoes	the quality remains optimal	10.1
Reporter	not notice influence of sounds coming from the environment, background noise and echoes	the quality remains optimal	10.2
Editor	the transcription quality is not lowered by regional accents	the quality remains optimal	11.1
Reporter	the transcription quality is not lowered by regional accents	the quality remains optimal	11.2
Editor	the transcription quality is not lowered by fast or slow speakers	the quality remains optimal	11.3
Reporter	the transcription quality is not lowered by fast or slow speakers	the quality remains optimal	11.4
Editor	the transcription quality is not lowered by the style of speech of 134 recurring members of parliament and ministers and external speakers	the quality remains optimal	11.5

Reporter	the transcription quality is not lowered by the style of speech of 134 recurring members of parliament and ministers and external speakers	the quality remains optimal	11.6
Editor and Reporter	get grammatically correct sentences	I can work faster and more efficiently	12
Editor and Reporter	get semantically correct sentences	I can work faster and more efficiently	13.1
Editor and Reporter	get semantically correct sentences with suggested alternatives	I can work faster and more efficiently	13.2
Reporter	get a logical and structured summary of the commission sessions	I can work faster and more efficiently	14
Editor	introduce a text of the corresponding speech into the system that is generated up front	the quality of speech-to-text goes up	15.1
Reporter	introduce a text of the corresponding speech into the system that is generated up front	the quality of speech-to-text goes up	15.2
Editor	count on a system where accuracy has priority over speed	I have less work correcting errors	16.1
Reporter	count on a system where accuracy has priority over speed	I have less work correcting errors	16.2
Editor	get a transcription integrated into my current workflow and existing applications	it is user friendly	17.1
Reporter	get a transcription integrated into my current workflow and existing applications	It is user friendly	17.2
Editor	count on the system to correctly handle with multiple members of parliament interrupting each other	I have less work correcting errors	18
Editor	count on the system to correctly handle redundant repetitions... in function of my own settings	I have less work correcting errors	19
Editor	count on the system to correctly handle hesitations, stuttering, expletives...	I have less work correcting errors	20

### 3 Synthesis: composing an RFP

#### 3.1 Phases and objectives of the commercial procurement

To meet the ambitions of this project, the necessary budget needs to be allocated and a lot of effort needs to be made. During the first phase the focus should be on a precommercial project. During this precommercial project a relevant prototype is built with limited resources to reduce the biggest sources of risk or even annihilate them. After the precommercial phase the Flemish Parliament can select the preferred functionalities and move on to phase 2 in which a common RFP is written and an integrated solution is developed. Finally, there is an option to move on to phase 3 in which innovative procurement of research is done.

Even after a precommercial procurement project, it is advised to build the integrated solution with a step by step approach. The final objective is to be obtained by developing incremental reachable improvements.

In the next section, the different phases are described and the corresponding use-cases and functionalities are summed up.

#### 3.2 Priorities of an integrated solution

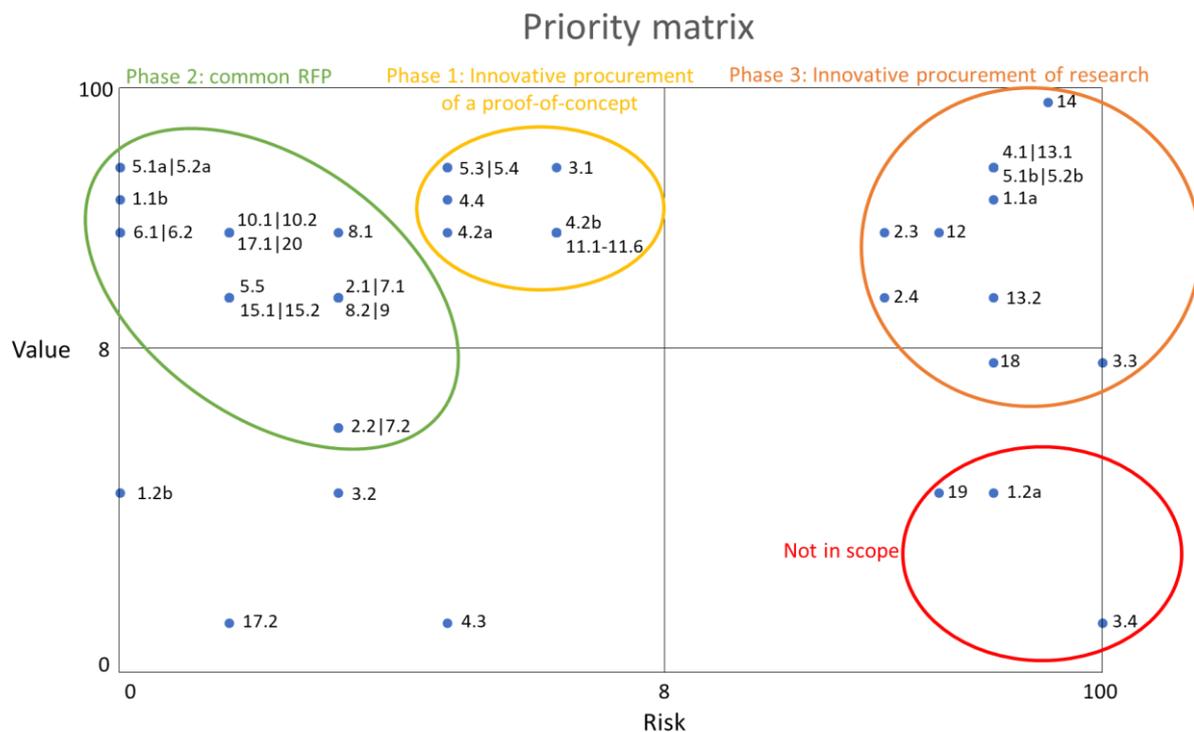


Figure 1 Prioritizing the functionalities

Figure 1 shows the priorities of the challenges and applications in a so called priority matrix of value vs risk. The vertical axis represents the value or score of the innovation potential. Scoring this innovation potential is done by showing the use-cases to end users via planning poker. The horizontal axis represents the technological risk, which is scored by showing the use-cases to industry players and knowledge institutes, also via planning poker.

Within this matrix one typically identifies four zones. On the top left are the critical functionalities of the final application that have high value for the end users but low implementation risks. On the top right are the elements that have high user value but also contain a high technological risk. On the bottom left are the details of the application, with low value and low risk. The final quadrant on the bottom right has low value for the end users and contains high technological risks. These functionalities are not important at all for the end product.

The highest priorities on the roadmap of a prototype are typically formed by the elements with a high added value for the end user and with a risk that is reducible on the short term. The higher the risk of a functionality, the harder to reduce it sufficiently on short notice. Depending on the available resources and spread of the functionalities on the grid, a minimal scoping of the prototype can be done.

### **3.2.1 Phase 1: innovative procurement of a proof-of-concept**

During the precommercial phase, the innovative procurement of a proof-of-concept, the scope needs to be made crystal clear in function of the user-groups to concretely realize the added value. At the same time, the technological and implementation risks need to be reduced.

During a first phase a precommercial procurement is launched that focusses on realizing the use-cases that can be supported by the existing technology but require a certain effort. After training and development of the prototype, evaluation of which functionalities to include in the final RFP can be done. In conclusion, the essence is reducing the risk of these use-cases.

The use-cases included are those with a high added value and a risk score between 3 and 8. Concretely the functionalities are efficiently introducing new speakers and corresponding speaker recognition, using a lexicon and handling different accents and styles of speech of the 124 members of parliament and a maximum of 11 ministers. This includes use-cases: 3.1, 4.2a, 4.2b, 4.4, 5.3, 5.4, 11.1-11.6.

In order to create a functioning whole, use-case 1.1, the transcription of audio to text in Word, needs to be included in the prototype as central functionality.

### **3.2.2 Phase 2: common RFP**

During the second phase an RFP is written in which both the use-cases with a low risk as well as the use-cases from phase 1 of which the risk is sufficiently reduced are included. When procuring the focus is on evaluating, integrating and deploying the application.

The use-cases involved are those with a high added value and a risk score lower than 3, complemented with the successfully developed use-cases of phase 1. The functionalities included are: a visible measure of accuracy of words or phrases, using editorial guidelines on word level, linking text with the corresponding audio-fragment, live editing of the transcription, robustness of the system, subtitling videos on the website, the influence of background noise, inputting a preconfigured text of the speech, integration in the existing workflow and handling hesitations and stuttering. This includes use-cases: 1.1b, 2.1, 2.2, 5.1a, 5.2a, 5.5, 6.1, 6.2, 7.1, 7.2, 8.1, 8.2, 9, 10.1, 10.2, 15.1, 15.2, 17.1, 20.

### **3.2.3 Phase 3: innovative procurement of research**

In a third and final phase research can be procured or supported. The essence is reducing the risk in the long term. This is only done for use-cases with a sufficiently high value from a user perspective. The archives with audio and corresponding reports can be of importance as training material of the research institutes.

Almost all use-cases included contain a component in which the semantics of speech play a vital role. These include, live recognition of speakers, recognizing the role of the chairman, grammatically and semantically correct sentences, structured and summarized reports and deleting needless repetitions.

The following use-cases are already included in phase 1 and 2, but can still be improved qualitatively through training and research: general transcription of speech to text and handling editorial guidelines (mainly those involving writing style and restructuring of sentences).

Phase 3 includes the following use-cases: 1.1a, 2.3, 2.4, 3.3, 4.1, 5.1b, 5.2b, 12, 13.1, 13.2, 14, 18.