

Lwazi II Final Report

Increasing the impact of speech technologies in South Africa

Compiled by the **Council for Scientific and Industrial Research (CSIR)**

February 2013

Contributors (CSIR Meraka Institute)

Dr Karen Calteaux

Dr Febe de Wet

Ms Carmen Moors

Mr Daniel van Niekerk

Mr Bryan McAlister

Ms Aditi Sharma Grover

Ms Tebogo Reid

Contributors (NWU Vaal Triangle Campus)

Prof Marelle Davel

Prof Etienne Barnard

Dr Charl van Heerden

Acknowledgements

This report was compiled under the auspices of the Council for Scientific and Industrial Research (CSIR) for the purpose of reporting on the Lwazi II project to the Department of Arts and Culture (DAC). A special word of thanks goes to all members of the Human Language Technology research group (current and past) who contributed to the project in some way.

Thanks are due to:

- Department of Arts and Culture
- Human Language Technology Research Group, CSIR Meraka Institute
- North-West University, Vaal Triangle Campus
- North-West University, Potchefstroom Campus
- Department of Basic Education, National School Nutrition Unit
- Thusong Service Centres (Bushbuckridge, Musina and Sterkspruit)
- Senqu Municipality
- Afrivet Training Services
- Kokotla Junior Secondary School
- Ntsako Junior Secondary School
- Inkonjane Middle School
- Thuto Pele Junior Secondary School
- Nqobangolwazi Junior Secondary School
- Reggie Masuku High School
- Mabusa Besala Secondary School
- Mthombo Secondary School
- Tshiwela Secondary School
- Marimane Secondary School
- Cedar Secondary School
- Ikeketseng Secondary School
- Kwezilomso Secondary School
- Cowen Secondary School
- Job Access

Report and marketing materials printed by MinitPrint

Executive summary

From 2010-2012 the Human Language Technology (HLT) research group of the CSIR Meraka Institute received funding from the Department of Arts and Culture (DAC) for a project entitled **Lwazi II: Increasing the impact of speech technologies in South Africa**. This is the final research report on this project.

The Lwazi II project followed on from Lwazi I and involved further research and development work on speech technologies for resource-scarce languages, mainly the official languages of South Africa. The project was divided into two subprojects – **Services** and **Technologies** – and the report has been structured accordingly.

The **Services subproject** involved the design, development, piloting, deployment, monitoring and evaluation of 11 services in domains relating to social services, education, health and agriculture. Two services were deployed for extended periods and one of these was hosted externally. Both deployed services involved the design, development and deployment of interactive voice response (IVR) systems linked to web interfaces for managing the incoming data. Transcriber interfaces enabled human transcription and translation of the inbound information. An outbound application was deployed to feed analysed and interpreted data back to users (manual analysis), thereby creating an information ecosystem. Business modelling for voice-based services as well as research into the business drivers for voice-based services and their implications for the design of such services complemented the technical work done in this subproject.

The **Technologies** subproject involved research and development work on text-to-speech (TTS) systems, automatic speech recognition (ASR), and the telephony platform used to deploy such technologies as well as the services described above. The TTS work focused on developing systems of improved quality that are demonstrably more natural sounding. These were achieved through Hidden Markov Models (HMMs) rather than unit-selection synthesis, as well as work on acoustic modelling and signal processing. Larger speech corpora with more prosodic variation were collected and more advanced natural language processing (NLP) modules were implemented in the TTS text-analysis front-end. The ASR work focused on improving the recognition accuracy for larger vocabularies. This was achieved by developing optimisation tools needed and incrementally optimising phone recognition starting with a limited number of languages and increasing these, and the levels of task difficulty, annually. The ASR research culminated in the development of systems that are able to achieve average accuracies of greater than 80% for vocabularies of 100-200 words in all 11 official languages. Work was also done on developing a speech processing pipeline to support the speech processing goals of the project and replace the previous proprietary software with a system which is openly available. A content management system (CMS) was developed to support quick and easy creation of speech-enabled applications for the project. The CMS manages content using a language-independent, distributed, web-

based management and maintenance system and provides the ability to administer and configure IVR and ASR systems with a single unified tool and language-independent configuration files. Lastly, work was done on hosting the IVR services externally, in order to enable and enhance scaling and maintenance of the services.

The report also features discussions on the evaluation of the above services and technologies with significant attention given to analyses of the call logs pertaining to the deployed IVRs as well as user experience evaluations relating to these services. Technology release is discussed in terms of the release and transfer of resources, systems and software, and services.

In addition to the technical outputs of the project, the **impact** of the project can be measured in terms of its other deliverables. These include 35 scientific outputs ranging from local and international conference papers to a book chapter; 13 students who worked on aspects of the project as part of their PhD and Master's studies; and skills development for a project team consisting of 29 members throughout its life-cycle. Abstracts of all the publications emanating from the project have been included in the report.

Besides these significant technical outputs, meaningful insight into the South African information technology landscape and potential end-user markets for voice-based services were also obtained:

- Our understanding of requirements analysis, application selection, scaling as well as monitoring and evaluation improved significantly during the Lwazi II project. Nevertheless, uptake of the services deployed remains a challenge. The development of an information ecosystem which ensures that users experience the benefits of using these services seems to be key to ensuring their sustainability, as well as to their deployment in commercial offerings.
- The speech resources and technologies that were developed during the Lwazi II project have started to reach levels of maturity which enable their deployment in customised small-scale commercial offerings. While this should be explored, research and development activities should also continue, in order to refine these technologies and improve their robustness and scalability.

Acronyms

API – Application programming interface
ASR – Automatic speech recognition
ATS – Afrivet Training Services
CDW – Community development worker
CHPC – Centre for High-Performance Computing, Meraka Institute, CSIR
CMN – Cepstral mean normalisation
CMS – Content management system
CMVN – Cepstral mean and variance normalisation
DAC – Department of Arts and Culture
DAFF – Department of Agriculture, Forestry and Fisheries
DBE – Department of Basic Education
DEA – Department of Environmental Affairs
G2P – Grapheme-to-phoneme
GSM – Global system for mobile communications
GUI – Graphical user interface
HCD – Human capital development
HLT – Human language technology
HLT RG – HLT research group, CSIR Meraka Institute
HMM – Hidden Markov Model
HTK – Hidden Markov Model Toolkit
HTS – HMM-based speech synthesis system
HTTP – Hypertext transfer protocol
IVR – Interactive voice response
Matlosana CAG – Matlosana Community Advisory Group
M&E – Monitoring and evaluations
MFCC – Mel-frequency cepstral coefficients
NCHLT – National Centre for Human Language Technology
NLP – Natural language processing
NSNP – National School Nutrition Programme
NWU – North-West University
POS – Part-of-speech
SDS – Spoken dialog system
SMS – Short message service
TSC – Thusong Service Centre
TTMP – Technology transfer master plan
TTS – Text-to-speech
UL – University of Limpopo
VUI – Voice user interface

Content

Executive summary	3
Acronyms.....	5
Content.....	6
List of Tables.....	12
List of Figures.....	13
1. Project overview.....	16
1.1 Context	16
1.2 Lwazi II project goals	16
1.3 Broad overview of planned deliverables.....	17
1.3.1 Subproject 1: Services	17
1.3.2 Subproject 2: Technologies	17
1.4 Project team	18
Subproject – Services	22
2. Services context and goals	24
3. Services overview	24
4. Community development workers (CDW) service	26
4.1 Background.....	26
4.2 Piloting.....	26
4.2.1 Casteel Thusong Service Centre	27
4.2.2 Madumbo Thusong Service Centre	27
5. Enhancement of the Senqu Municipality CDW service	28
5.1 Background.....	28
5.2 Piloting.....	28
5.2.1 Senqu Municipality	28
5.2.2 Sterkspruit Thusong Service Centre	29
5.3 Evaluation of the CDW service	29
6. Job search service.....	29
6.1 Background.....	29
6.2 Design	30
6.3 Development	30
6.4 Piloting.....	30
6.5 Evaluation	31
7. DBE services.....	31
7.1 Background.....	31

7.2	Design	32
7.2.1	Learner feedback service.....	32
7.2.2	Coordinator reporting service	34
7.2.3	Web monitoring interface	34
7.3	Development	34
7.4	Deployment	35
7.4.1	Learner feedback service.....	39
7.4.1.1	Introduction.....	39
7.4.1.2	Methodology	39
7.4.1.3	Observations.....	39
7.4.1.4	Challenges.....	42
7.4.1.5	Opportunities	43
7.4.2	Coordinator reporting service	44
7.4.2.1	Introduction.....	44
7.4.2.2	Methodology	44
7.4.2.3	Observation and comments	45
7.4.3	Involvement of district NSNP officials	45
7.4.3.1	Introduction.....	45
7.4.3.2	Methodology	45
7.4.3.3	Questions.....	46
7.4.4	SMS Report reminder service	46
7.4.5	Conclusion	47
7.5	Call log analysis.....	47
7.5.1	Learner feedback service.....	47
7.5.1.1	Call volume	48
7.5.1.2	Call duration	53
7.5.1.3	Usage frequency analysis	53
7.5.1.4	Language choice	54
7.5.1.5	Final call state	64
7.5.1.6	Invalid entries	67
7.5.1.7	Timeouts	68
7.5.2	Coordinator reporting service	69
7.5.2.1	Call volume	69
7.5.2.2	Call duration	74
7.5.2.3	Final call state	74

7.5.2.4	Invalid entries	77
7.5.2.5	Timeouts	78
7.5.3	Discussion	79
7.6	User experience evaluation.....	80
7.6.1	Learner feedback service evaluation.....	80
7.6.1.1	Background.....	80
7.6.1.2	Objectives	81
7.6.1.3	Research methodology.....	82
7.6.1.4	Findings and recommendations	83
7.6.2	Coordinator reporting service evaluation	89
7.6.2.1	Objectives	89
7.6.2.2	Research methodology.....	89
7.6.2.3	Findings and recommendations	90
7.6.3	Web monitoring interface evaluation	90
7.6.3.1	Objectives	91
7.6.3.2	Research methodology.....	91
7.6.3.3	Findings and recommendations	91
7.6.4	Conclusions.....	91
8.	Afrivet services	92
8.1	Background.....	92
8.2	Design	93
8.2.1	Afrivet System Overview	93
8.2.2	V-Plan Vet line	94
8.2.3	Dip tank line.....	95
8.2.4	Web management interfaces	95
8.2.5	Outbound alerting service	96
8.3	Development	97
8.4	Piloting.....	98
8.4.1	V-Plan Vet line	98
8.4.2	Dip tank line.....	100
8.4.2.1	Methodology	101
8.4.2.2	Observations (version 1.0 piloting)	102
8.4.3	Questions.....	103
8.4.4	Challenges.....	103
8.5	Evaluation	104

8.5.1	V-Plan vet line.....	104
8.5.1.1	Process.....	104
8.5.1.2	Data analysis.....	104
8.5.1.3	Recommendations.....	105
8.5.2	Dip tank line.....	106
8.5.2.1	Process.....	106
8.5.2.2	Results	106
8.5.2.3	Challenges.....	107
8.5.2.4	Recommendations.....	107
8.5.3	Conclusions.....	107
8.6	Call log analysis.....	107
8.6.1	Dip tank line.....	108
8.6.1.1	Call volume	108
8.6.1.2	Call duration	109
8.6.1.3	Usage frequency analysis	109
8.6.1.4	Final call state	110
8.6.1.5	Invalid entries	112
8.6.1.6	Timeouts.....	112
8.6.1.7	Barge-in	113
8.6.2	V-plan vet line.....	113
8.6.2.1	First pilot: Call log analysis.....	114
8.6.2.2	Second launch: Call log analysis	118
8.6.2.3	Call volume and reporting frequency.....	118
8.6.3	Discussion	120
9.	Services-related HCD and knowledge generation.....	121
10.	Services conclusions	121
	Subproject – Technologies	126
11.	Text-to-speech (TTS) conversion	128
11.1	Introduction.....	128
11.2	Activities	128
11.2.1	Research and development for HMM-based speech synthesis	128
11.2.1.1	Acoustic modelling and synthesis.....	128
11.2.1.2	Pitch modelling.....	129
11.2.1.3	Code-switching	129
11.2.1.4	Speech corpus development	130

11.2.2	Natural language processing	132
11.2.2.1	Text tokenisation and normalisation.....	132
11.2.2.2	Part-of-speech tagging	132
11.2.2.3	Lexical stress and tone.....	132
11.2.2.4	Morphological analysis.....	133
11.2.2.5	Higher-level syntactic and semantic analysis	133
11.3	Results	133
11.4	Summary of outputs.....	134
11.5	Software	134
11.6	Data	135
12.	Automatic speech recognition (ASR).....	135
12.1	Context and goals.....	135
12.2	Main research themes.....	136
12.2.1	Modelling code-switched speech.....	136
12.2.2	Improved pronunciation modelling.....	137
12.2.3	Data combination and sharing	137
12.2.4	Trajectory modelling.....	138
12.3	ASR HCD.....	138
12.4	ASR system optimisation	139
12.4.1	Optimised phoneme recognition.....	139
12.4.1.1	Evaluation protocol	140
12.4.1.2	ASR system development toolkit.....	140
12.4.1.3	Optimisation techniques	141
12.4.1.4	Results	142
12.5	Conclusion	144
13.	Platform development	144
13.1	Speech processing pipeline	144
13.1.1	Description.....	144
13.1.2	Requirements	145
13.1.3	Design	145
13.1.4	Implementation report.....	147
13.1.5	Testing	147
13.1.6	Discussion	148
13.1.7	Looking forward.....	149
13.2	Content management system	149

13.2.1	Description.....	149
13.2.2	Requirements	149
13.2.3	Design	149
13.2.4	Testing	150
13.2.5	Progress	150
13.3	Telephony platform improvements	151
13.3.1	Platform run time testing	151
13.3.2	Externally hosted infrastructure.....	151
13.3.2.1	Description.....	151
13.3.2.2	Motivation	152
13.3.2.3	Solution design	152
13.3.2.4	Testing	153
14.	Technology evaluation	153
15.	Technology release.....	155
15.1	High-level technology overview	155
15.2	Lwazi II technologies to be transferred	156
15.2.1	Transfer of resources.....	156
15.2.2	Transfer of systems and software	157
15.2.3	Transfer of applications.....	158
15.2.4	Licensing	158
16.	Impact.....	159
16.1	Societal impact	160
16.1.1	In-depth requirements analysis sessions	160
16.1.2	Rapid prototyping and testing.....	160
16.1.3	Understanding the business drivers.....	160
16.1.4	Business modelling	161
16.2	Human capital development.....	161
16.3	Scientific impact	162
16.3.1	Overview of scientific publications.....	162
16.3.2	Abstracts – ASR publications	170
16.3.3	Abstracts – TTS publications.....	177
16.3.4	Abstracts – Speech applications	180
17.	Summary.....	186
18.	References.....	188

List of Tables

Table 1: Project team members	19
Table 2: Summary of services deliverables per project year	26
Table 3: List of schools from all 3 phases	37
Table 4: Total call distribution per month per province.....	48
Table 5: Total call distribution per month per school	50
Table 6: Average calls per week	52
Table 7: Usage frequency analysis.....	54
Table 8: Language choice distribution per school per province.....	56
Table 9: Gauteng – Language choice distribution per school	58
Table 10: North West – Language choice distribution per school	59
Table 11: Mpumalanga – Language choice distribution per school.....	60
Table 12: Eastern Cape – Language choice distribution per school.....	61
Table 13: Free State – Language choice distribution per school.....	62
Table 14: Limpopo – Language choice distribution per school	63
Table 15: Language choice distribution per province	63
Table 16: Distribution of final call states	65
Table 17: Final state distribution per school	66
Table 18: Invalid entries – All schools.....	67
Table 19: Timeouts – All schools	68
Table 20: Call distribution per school per month (Coordinator reporting service).....	70
Table 21: Average calls per week	73
Table 22: Final call state per school.....	76
Table 23: Final call state – "Unknown" schools.....	77
Table 24: Invalid entries per school.....	77
Table 25: Timeouts per school	78
Table 26: Evaluations of the learner feedback service.....	80
Table 27: Average calls per week	109
Table 28: Usage frequency distribution	110
Table 29: Distribution of final call states	111
Table 30: Invalid entries	112
Table 31: Timeouts	113
Table 32: Barge-in distribution	113
Table 33: Registered users and calls received.....	115
Table 34: Call success rate	115

Table 35: Usage frequency distribution – Vets.....	116
Table 36: Call duration – Single reports	117
Table 37: Call duration – Multiple reports	117
Table 38: Total call distribution per month	118
Table 39: Average calls per week	119
Table 40: Distribution of final call states	120
Table 41: Consolidated view of the Lwazi II services, piloting and evaluation.....	122
Table 42: Speech corpora recorded during the project	131
Table 43: Perceptual test results	134
Table 44: ASR results for all 11 languages.....	142
Table 45: ASR results for all 11 languages on clean subset of test data only	143
Table 46: ASR results for systems with improved pronunciation modelling	143
Table 47: ASR research and development goals	144
Table 48: 200-term recognition accuracy for various decoders and feature sets	148
Table 49: Technology evaluation criteria	154
Table 50: Resource release.....	157
Table 51: Software release	158
Table 52: Licensing.....	159
Table 53: PhD studies	161
Table 54: Master’s studies.....	162
Table 55: Publications overview	162
Table 56: ASR publications	164
Table 57: TTS publications.....	166
Table 58: Speech applications publications.....	167
Table 59: Income and expenditure per project year.....	Error! Bookmark not defined.
Table 60: Income and expenditure over the project period (Jan 2010 – Dec 2012).....	Error! Bookmark not defined.
Table 61: Lwazi deliverables.....	186

List of Figures

Figure 1: Subproject 1 – Services.....	17
Figure 2: Subproject 2 – Technologies.....	18
Figure 3: Original Learner feedback service pamphlet.....	33
Figure 4: Revised Learner feedback service pamphlet.....	33
Figure 5: Web interface: Learner feedback service.....	35

Figure 6: Learners at Iketsetseng Secondary School	41
Figure 7: Learners at Cedar Secondary School	42
Figure 8: Piloting at Mthombo Senior Secondary School.....	44
Figure 9: Total call distribution per month.....	48
Figure 10: Call duration distribution.....	53
Figure 11: Language distribution in all schools and provinces.....	57
Figure 12: Language choice – Gauteng.....	57
Figure 13: Language choice – North West.....	58
Figure 14: Language choice – Mpumalanga	59
Figure 15: Language choice – Eastern Cape	60
Figure 16: Language choice – Free State	61
Figure 17: Language choice – Limpopo	62
Figure 18: Final call state for all schools.....	64
Figure 19: Total no. of calls across all schools.....	72
Figure 20: Call duration distribution.....	74
Figure 21: Final call state	75
Figure 22: Overview of PAHC role players.....	92
Figure 23: Afrivet system architecture	93
Figure 24: A sample V-plan Vet line interaction.....	94
Figure 25: Web transcription interface	96
Figure 26: Outbound alerting service (beta)	97
Figure 27: Front page of V-Plan Vet line brochure	99
Figure 28: Back page of V-Plan Vet line brochure	99
Figure 29: Dip tank attendants attending the pilot.....	101
Figure 30: Dip tank attendants testing the system	102
Figure 31: Dipping products supplied by DAFF and DEA	103
Figure 32: Total call distribution per month.....	108
Figure 33: Call duration distribution.....	109
Figure 34: Final call states	111
Figure 35: Total call distribution per day – Overall	114
Figure 36: Total call distribution per day – Vets.....	115
Figure 37: Call time distribution – Vets	116
Figure 38: Call duration distribution.....	119
Figure 39: Final call states	120
Figure 40: Speech processing pipeline design.....	147

Figure 41: Content management system design.....	150
Figure 42: DBE Learner IVR application up and down time	151
Figure 43: External hosting configuration	153
Figure 44: High-level technology overview	156

1. Project overview

1.1 Context

From 2006 to 2009, the Department of Arts and Culture (DAC) sponsored the **Lwazi¹** project, in which various basic speech technologies for all the official South African languages were developed, and a telephone-based information system for community development workers (CDWs) was piloted at small-scale. To increase and expand the impact of human language technologies in South Africa, DAC sponsored the **Lwazi II** project from 2010 to 2012. This document is the final research report for the Lwazi II project.

1.2 Lwazi II project goals

Lwazi II aimed to elaborate on the work that was done in Lwazi I, thereby ensuring a positive growing trend in the development of multilingual speech technologies that would have a positive impact on the lives of South Africans.

The project had two main goals:

- To pilot and deploy a number of additional telephone-based services, in order to demonstrate the impact of multilingual telephone-based services in meeting the actual needs of users.
- To improve, expand and enrich various speech technologies, in order to ensure greater uptake of these technologies in customer-facing services.

To this end, the proposed project consisted of two closely-related **subprojects**, phased over a period of three years. **Subproject 1: Services** ensured the piloting and deployment of various telephone-based information systems, while **Subproject 2: Technologies** focused on the enhancement of existing technologies for all official languages.

Subproject 1 specifically focused on determining –

- how impact can be increased through telephone-based services; and
- how such services can be deployed effectively and efficiently.

In Subproject 2, the technology enhancements aimed to –

- benefit the extended applications that would be deployed during the project, where relevant;
- improve the quality of each of the speech-technology components; and

¹ Referred to as Lwazi I in this report in order to distinguish between it and the Lwazi II project.

- increase the utility of these components for researchers and developers in other domains (for example, education, books for visually impaired users, and information kiosks).

1.3 Broad overview of planned deliverables

1.3.1 Subproject 1: Services

The CSIR was contracted to deliver the following for this subproject:

1. Piloting of the community development workers' (CDW) telephone-based service at two additional sites, thereby covering all eleven South African languages with the CDW telephone-based service (including the Lwazi I pilots).
2. Development and piloting of four additional telephone-based services, piloted at selected sites for selected languages in selected domains.
3. Deployment of two telephone-based services (which might be the same as the ones piloted), deployed fully until completion of the contract at selected sites for selected languages in selected domains.
4. Development of a technology transfer master plan to direct the deployment (roll-out) of similar future telephony services by government (or other implementation agencies).

Figure 1 provides an overview of the proposed services and their domains.

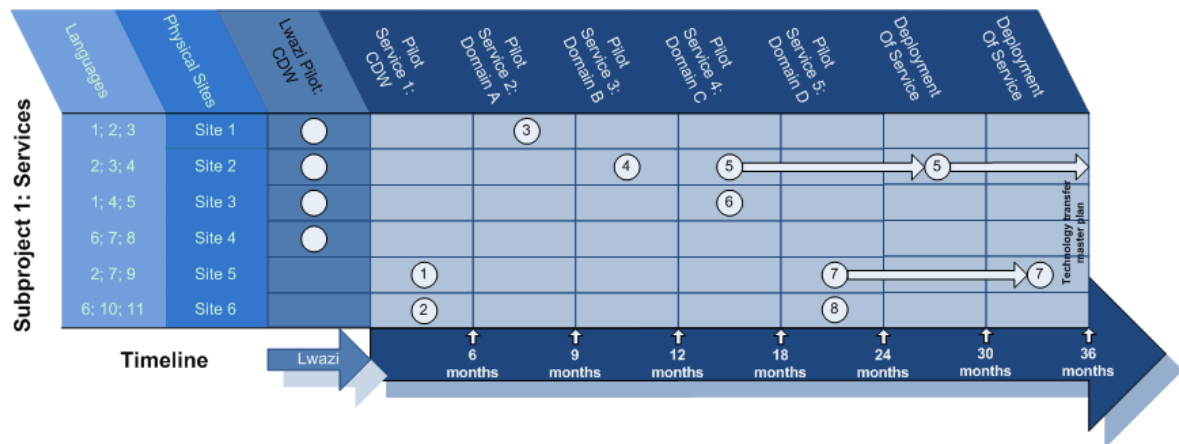


Figure 1: Subproject 1 – Services

1.3.2 Subproject 2: Technologies

The CSIR was contracted to deliver the following for Subproject 2:

1. Development of a distributed, web-based management and maintenance system, in order to allow for the large-scale distribution of highly localised content.

2. Closer integration of the speech recognition engine with the IVR system, in order to support more concurrent sessions and more diverse services running in parallel on a single system.
3. Experimental investigation of integration with alternative telecommunication infrastructures, in order to explore novel opportunities for the use of our platform with emerging infrastructures.
4. For all official South African languages, ASR systems achieving accuracies of greater than 80% for vocabularies of 100-200 words.
5. Improved and extended linguistic resources, including larger speech corpora and more accurate acoustic models, phone sets and pronunciation dictionaries, as required to reach the goal in 4. above.
6. For all official South African languages, TTS systems that are demonstrably more natural-sounding than the initial voices developed as part of the Lwazi I project.
7. Additional linguistic resources, including annotated text corpora, as required for reaching the goal in 6. above.

Figure 2 provides an overview of the technologies and their delivery dates.

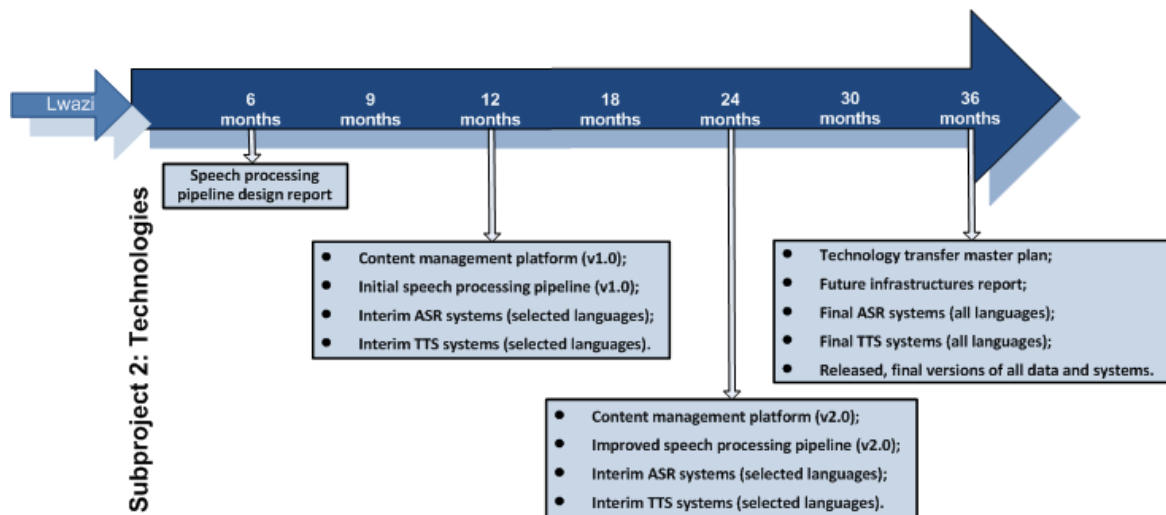


Figure 2: Subproject 2 – Technologies

1.4 Project team

As with any research project, the research team changed over the course of the project. The table below provides an overview of the project team over the life of the project. The names of members who left the team have been struck out, while the names of members joining the team after 1 January 2010 have been indicated in italics.

Table 1: Project team members

Name	Team	Position	Status
Aby Louw	TTS	Researcher	Ongoing
Aditi Sharma Grover	Design	Researcher/Technical team leader	Ongoing
<i>Akhona Samson</i>	<i>M&E</i>	<i>Intern</i>	<i>Joined: July 2011</i>
Alta de Waal	ASR, TTS	Research Group Leader	Resigned: May 2012
Bryan McAlister	Software development	Developer	Ongoing
<i>Carmen Moors</i>	<i>Project management</i>	<i>Project Manager</i>	<i>Joined: April 2011</i>
Charl van Heerden	ASR	Student	Resigned: Dec 2011
Daniel van Niekerk	TTS	Researcher/Technical team leader	Resigned at project end
Georg Schlünz	TTS	Researcher	Ongoing
Gerhard van Huyssteen	Design	Research Group Leader/ Consultant	Resigned: Oct 2010/ Ongoing
Jaco Badenhorst	ASR	Researcher	Ongoing
Jama Ndwe	Design	Researcher	Resigned: April 2011
Karen Calteaux	Design, M&E	Research Group Leader	Ongoing
<i>Laurette Pretorius</i>	<i>TTS</i>	<i>Student</i>	<i>Ongoing</i>
Marelle Davel	ASR	Researcher	Resigned: June 2011
Mixo Shibus	Software development	Developer	Resigned: Jan 2011
Motshana Phohole	M&E	Researcher	Resigned: April 2012
Mpho Kgampe	M&E	Student/Researcher	Ongoing
Mpho Raborife	TTS	Researcher	Resigned: June 2012
Neil Kleynhans	ASR	Student/Researcher	Ongoing
Nina Titmus	ASR, TTS	Project Manager	Ongoing
Olwethu Qwabe	Design	Researcher	Ongoing
Onkgopotse Molefe	M&E	Researcher	Re-assigned: June 2011
Richard Carlson	Software development	Consultant	Resigned: Oct 2012
Tebogo Gumede	M&E	Researcher/Technical team leader	Ongoing
Thipe Modipa	ASR	Researcher	Ongoing

Name	Team	Position	Status
Tshepo Moganedi	Software development	Developer	Ongoing
<i>Willem Basson</i>	<i>ASR</i>	<i>Student</i>	<i>Resigned at project end</i>
Zachary Battles	Design	Researcher	Resigned: June 2012

Subproject – Services

2. Services context and goals

Lwazi I focused on developing and piloting a telephone-based, speech driven information service for improving government service delivery. The main goals of the Lwazi II services subproject were to increase the impact of telephone-based applications² in South Africa by –

- developing more than one type of service³;
- piloting these services in more than one domain⁴ for selected languages; and
- deploying two of the services over a longer period of time.

The following secondary goals were also important:

- Testing speech technologies by means of real-world applications
- Contributing to new knowledge with regard to –
 - the selection of applications aimed at identifying needs which can be best addressed by means of telephone-based technologies;
 - designing voice user interfaces for the South African emerging market;
 - rapid application development in the context of interactive voice response systems (IVRs); and
 - deploying telephone-based services within the South African context.
- Raising awareness of the benefits of telephone-based services for sharing, retrieving and accessing information and improving service delivery.
- Human capital development.

3. Services overview

Over the course of the Lwazi II project, 11 services were piloted. Nine of these required a complete development cycle – from selection, design, and development to piloting, and evaluation. In 2012, two sets of services (the DBE and the Afrivet services) were selected for full-time deployment until the end of the project. A brief description of the various services follows:

- **Service 1 – CDW service pilots:** Two additional pilots of the CDW service were required in order to ensure that the CDW application had been piloted in all 11 official languages.

² An application is a computer software programme designed to help the user perform singular or multiple related specific tasks in order to solve problems in the real world.

³ Applications designed to meet the needs of a client or the research agenda of the HLT research group.

⁴ Areas of specialisation such as education, social services, health and so on.

- **Service 2 – Enhancement of the Senqu TSC CDW service:** SMS functionality was added to the Senqu TSC CDW IVR due to poor network coverage resulting in poor quality and intermittent availability of the IVR. A new generation isiXhosa TTS voice was also added as there had been reports of users experiencing difficulty in understanding the Lwazi I TTS voice.
- **Service 3 – DBE, NSNP SMS-based report reminder service:** A web interface is used to set up SMS reminders to sets of recipients to remind them of due dates and/or share information. This saves time and renders a more efficient and effective audit trail.
- **Service 4 – DBE, NSNP Learner feedback service:** An IVR enables learners and their caregivers to make a free call to give feedback on meals received or not received; what the meals entailed; and the quality of the food provided (in terms of their like/dislike thereof).
- **Service 5 – DBE, NSNP Coordinator reporting service:** This service is an IVR which facilitates school coordinator reporting on services rendered. Since this reporting is done daily and over the phone, there is less paper work and recall discrepancies.
- **Service 6 – DBE NSNP Web-based monitoring service:** The information from the Coordinator and Learner IVRs is visualised on a web interface which also provides for information management such as transcription of audio messages, and downloading of reports.
- **Service 7 – Job search:** An IVR enables users who do not have internet access or other alternatives, to capture and verify a CV.
- **Service 8 – Afrivet V-plan vet line:** Disease surveillance information pertaining to livestock is obtained from rural private veterinarians by means of an IVR.
- **Service 9 – Afrivet Dip tank line:** Dip tank attendants (livestock owners) report on the condition of livestock and conditions at their dip tank by means of an IVR.
- **Service 10 – Afrivet Web-based monitoring service:** The information from the Vet line and Dip tank line IVRs is visualised on a web interface which also provides for information management such as transcription of audio messages, downloading of reports, and user registration.
- **Service 11 – Afrivet Outbound service:** A web interface enables the client to send outbound messages to defined sets of users via SMS or email, in one or more languages (manual translation), according to user preference.

Table 2 summarises the expected deliverables, outputs and target dates for the services developed during Lwazi II.

Table 2: Summary of services deliverables per project year

Timeframes	Goals	Outputs
Jan-Dec 2010	Additional piloting of CDW service to cover all official languages	Two CDW service pilots
	Design, develop and pilot 2 additional services (selected languages in selected domains)	Sterkspruit CDW enhancement
		DBE SMS service (piloted Jan 2011)
Jan-Dec 2011	Design, develop and pilot 4 additional services (selected languages in selected domains)	DBE learner service (phases 1 & 2)
		DBE coordinator service (phases 1 & 2)
		DBE Web monitoring service
		Job search service
Jan-Dec 2012	2 services deployed for remained of project	3 DBE services (phase 3)
		3 Afrivet services

In the sections which follow, the Lwazi II services will be discussed in terms of their background, design, development, piloting, and evaluation (as appropriate). Two types of evaluation were undertaken for the deployed services (the DBE and Afrivet services), namely user experience evaluations (by means of focus group and individual interviews) and usability/design evaluations (by means of call log analyses).

4. Community development workers (CDW) service

4.1 Background

During Lwazi I, the CDW service was piloted in seven official languages. Two additional pilots of the CDW service were required in order to ensure that all 11 official languages had been covered in the piloting of this service. No changes were made to the Lwazi I CDW design and no additional development work was required.

4.2 Piloting

The pilots took place at Casteel Thusong Service Centre (TSC), Bushbuckridge and Madimbo Thusong Service Centre, Musina. The languages covered were Xitsonga, Tshivenda, isiNdebele and siSwati. The pilot sites were carefully selected to ensure that the above languages would be targeted. The actual piloting was preceded by visits to the pilot sites to build rapport with the relevant contact persons and determine the suitability of the sites.

4.2.1 Casteel Thusong Service Centre

The same protocol was followed as for the previous CDW pilots (CSIR Meraka Institute, 2009).

Feedback from the workshop can be summarised as follows:

- The cost- and time-saving aspects of the service attracted the attention of the CDWs and are most likely to be the main drivers behind its uptake.
- Concerns about receiving messages not meant for them were clarified.
- Questions were raised on whether and how the community could use the service to communicate with the TSC manager.
- The system was easy to understand and use, and was well-received.

Increasing the uptake of the service would require a dedicated marketing campaign, repeated interaction with the potential users, and continuous encouragement to use the service.

4.2.2 Madumbo Thusong Service Centre

Due to the deep rural location selected for this pilot, a different protocol was followed to the other pilots. This involved obtaining permission from the headman to meet with the relevant departmental officials, having to make use of interpreters.

Feedback from the workshop can be summarised as follows:

- The TSC manager easily followed the instructions and understood the whole system as he was explaining it to the headmen.
- Although one of the older headmen did not seem to understand the system, the others did and were keen to use the system.
- The Lwazi team was thanked for bringing things that are “mahala” (free) and was requested to bring more things to this rural area.
- Low literacy levels led to most of the communication having to be done in Tshivenda.
- The new generation Tshivenda TTS voice was used. It is a drastic improvement on the Lwazi I TTS voices.
- The users suggested that the English numbers and times be removed and that Tshivenda numbers and times be used.
- Several questions were raised over the administration of the Lwazi system, such as whether –
 - it would be possible to change the recipients’ names in the system when the structures change;
 - Lwazi is affected by network problems;
 - the service was only being piloted in Limpopo or Madimbo;

- someone else would be able to add recipients to the system in the absence of the TSC manager;
- it was possible to leave a message on the system for someone to call one back, without that person incurring costs; and
- the service would be run over an extended period of time.

Challenges experienced at Madimbo:

- Internet access – the civic structures suggested that we register them separately; as separate structures that are from Madimbo and Dimboni villages. It later became evident that this would not be possible since they did not have internet access.
- Uptake – one person said that for the next few months the usage would be very low but as they got used to the system the usage would start increasing.
- Access to cell phones –the Lwazi team drove around Madimbo Village looking for people to test the system and went to Madimbo clinic. The team found that, unlike at the previous sites where Lwazi was piloted, people in this area did not have cell phones with which to test the system.
- Communication – Tshivenda is the one language that our team members do not understand, making it difficult to put across information. The TSC manager was interpreting but it was clear that not all the information was being translated.

5. Enhancement of the Senqu Municipality CDW service

5.1 Background

Based on a request from the Senqu Municipality, the CDW service of Lwazi I was enhanced through the addition of the capability to send SMSes (through the CDW website), as well as the addition of improved TTS for isiXhosa. The SMS feature allows CDW managers to have an option between using a voice (the CDW telephone service) or an SMS channel, to disseminate announcements to the CDWs and their communities. Similar to the voice channel the SMS feature allows the manager to send out the SMS in multiple languages.

5.2 Piloting

5.2.1 Senqu Municipality

The enhanced service was piloted with Tlotlisang Koena, the communications officer for the Senqu municipality. The new features which had been added to the Lwazi website were explained and she was shown how to use the new SMS service.

She was satisfied with the new service and hopeful that it would be used more regularly. The possibility of allowing her to add new users and to send messages to users who are not registered on the system was discussed. For the time being, it was decided that adding new

users to the system should be done by the Lwazi team, in order to manage the system effectively. This is, however, a matter which should be revisited in the future.

The team observed that Ms Koena had used the service that night to send a message about a meeting to 30 recipients.

5.2.2 Sterkspruit Thusong Service Centre

The enhanced application was piloted with Mr Mlindeli Sunduza, the centre manager. The added features of the SMS service were explained to him, as were the benefits of using both the SMS and the voice services. He was requested to use both the voice and SMS services as both are crucial to the improvement of human language technologies in South Africa.

The HLT team monitored the introduction of the SMS service along with the CDW voice application to see whether it will increase the use of the Lwazi service. The enhancement enabled the Lwazi team to explore users' preferences with regard to voice and text (in English and isiXhosa). This information will guide and inform future language technology research.

It is hoped that this service will enable us to establish the costs of running such services at local government level.

5.3 Evaluation of the CDW service

It was found that, despite initial excitement over the benefits of using the Lwazi CDW service, there has been very limited uptake. The exact reasons for this were not researched by the project team during Lwazi II, but discussions during the piloting suggest that these may relate to issues such as –

- the quality of the synthesised speech (as reported by the Senqu pilot site);
- limited awareness of the benefits of the service;
- limited awareness of its ease-of-use; and
- user perceptions relating to technology (including aspects such as a lack of trust in the service, concerns over privacy of information, and misunderstandings around who has/does not have the “right” to enter messages on the system).

6. Job search service

6.1 Background

This application aimed to develop a telephone-based, speech-driven interface to enable job seekers to place an abridged CV with a web-based recruitment agency, in order to become “visible” in the job market. This was a form-filling type application. Difficulties relating to the nature of the information to be captured, the typical user profile, and the user experience

were encountered during the design phase. The service was developed up to prototype level, tested with a small set of users (lab test) and a decision was made not to progress with its development as part of Lwazi II.

6.2 Design

The Job search service allows the user to capture (via audio recording) and review (via TTS) their CV telephonically. The call flow was designed with a stage-gated workflow process, viz. 1) record CV; 2) submit CV for transcription; 3) review the transcribed CV; and 4) submit the transcribed CV to Job Access (a partner recruitment agency – <http://www.jobaccess.co.za/>). The user gets a personalised user experience, since the system keeps track of their progress in the application (in terms of the above-mentioned stages), and only poses to the user relevant options with respect to the stage completed in their particular CV submission process.

The complexity of the call flow design was further increased by the detailed checks performed in each stage to ensure that the user has completed all the sections per stage, e.g. recording all the required aspects of the CV, and reviewing each transcribed audio segment of the CV.

6.3 Development

Voxeo, a web-based platform for developing telephony applications (<http://www.voxeo.com/>), was used to build two IVR systems, one to capture CV information and one to review a sample transcribed CV, read back to the user using the Lwazi TTS English voice.

6.4 Piloting

A telephonic experiment was conducted on 20 December 2011. The aim of this experiment was to obtain information from users concerning the use of a telephone to record and verify curriculum vitae (CV) information.

The four participants (two visually impaired, one mobility impaired, and one with a mental disorder) for this pilot were drafted from a pool of job seekers who are registered with Job Access. Participants were selected based on a mix of race, gender, disability and location. Interaction with participants was done telephonically as transport is a problem for persons with disabilities in South Africa.

The process for this experiment was as follows:

- The piloting team called the participant, introduced the project and explained the procedure for the test calls. The number for the “record CV line” was then given to the participant.
- The participant then called the IVR and recorded his/her CV.
- The piloting team called the participant after 10 minutes and explained the following part of the experiment. The participant was then given the “verify CV line” number to call.
- The participant called the second IVR and verified the sample CV which had been sent to him/her prior to the experiment.
- The piloting team then called the participant a third time and interviewed him/her on his/her experience of using both systems.

6.5 Evaluation

All participants completed the exercise and no real difficulties were encountered while using the IVR systems. One participant commented that the volume on prompts was too low, but none of the other participants found this to be a problem. Feedback received during the interviews was mainly positive and although a telephone is not considered a traditional tool to build a CV, preliminary results indicate that a telephone could be used to capture and review CV information – especially by users who 1) do not have any other means of capturing a CV; or 2) are blind. Various aspects of the implementation of such a service, including it being transcription-intensive, led to a decision not to develop the prototype further during Lwazi II.

7. DBE services

7.1 Background

Four services were developed for the National School Nutrition Programme (NSNP) of the Department of Basic Education (DBE), namely –

- **an SMS report reminder service** through which SMS reminders are sent to the provincial NSNP coordinators, to remind them of due dates for their reports and inform them of important issues which require their attention;
- **a Learner feedback service** which enables learners to give anonymous feedback on the quantity and quality of the food they receive through the NSNP;
- **a Coordinator reporting service** which enables the NSNP school coordinators to provide daily feedback on the meals served at their schools, including the number of learners fed; and

- a **Web monitoring service** which enables the NSNP staff at national, provincial and district level to monitor feedback from the learners and school coordinators and react to issues raised.

7.2 Design

The design activities focused on the design of two voice user interfaces (VUIs) – one for learners and one for school coordinators – as well as the design of two graphical user interfaces (GUIs) – one for the SMS report reminder service and one for the web monitoring interface.

The Learner feedback service design is particularly noteworthy here, as it entails developing a VUI for children. There has been little research on developing VUIs for this target audience to date. The multilingual nature of this service further complicated the prompt design. Consultation with experts on child behaviour was required during the prompt design process in order to ensure that the prompts would be appropriate for the cognitive development levels of the target audience.

7.2.1 Learner feedback service

The Learner feedback service consists of two parts: a multilingual telephone-based interactive voice response (IVR) line that allows learners to provide feedback on their school meals, and a web monitoring interface which provides the NSNP with up-to-date and timely information in both condensed and detailed formats across all the schools where the system has been piloted to date.

During the design of the service various factors such as language choice, input modality, and the type and length of the IVR interaction, required further investigation since the targeted user audience of the service was children (an audience where the use of IVR has not been researched extensively) (Sharma Grover *et al.*, 2012). The design process took a multi-perspective approach by involving various stakeholders such as learners, national-level NSNP officials, NSNP school coordinators, a child psychologist, and language practitioners. The call flow and prompt design was done in collaboration with a child psychologist to ensure the service was simple, clear and friendly for use by our target audience. An iterative design process ensured that the service addressed all the NSNP feeding monitoring measures (i.e. time of feeding, meal satisfaction, etc.) through relevantly posed questions in the IVR. The service also allowed learners to provide general audio messages on any issues related to the feeding at their school, which provided for richer qualitative data.

Extensive usability tests and focus group discussions were conducted, where learners experimented with four versions of the service. Specifically, we investigated the learners' choice of language with respect to English versus local languages, as well as input modality with respect to using speech input versus dual-tone multi-frequency (DTMF)/touchtone

input. The results were analysed extensively through quantitative and qualitative approaches and presented through an internationally peer-reviewed publication. Great care and effort was also made in crafting the system persona of “Mama Nandi” as trust and anonymity were found to be crucial factors for learners to use the service. The call flow design was also adapted to create different versions that cater for Gauteng and the other provinces, since Gauteng serves 2 meals (breakfast and lunch) and other provinces serve 1 meal a day.



Figure 3: Original Learner feedback service pamphlet

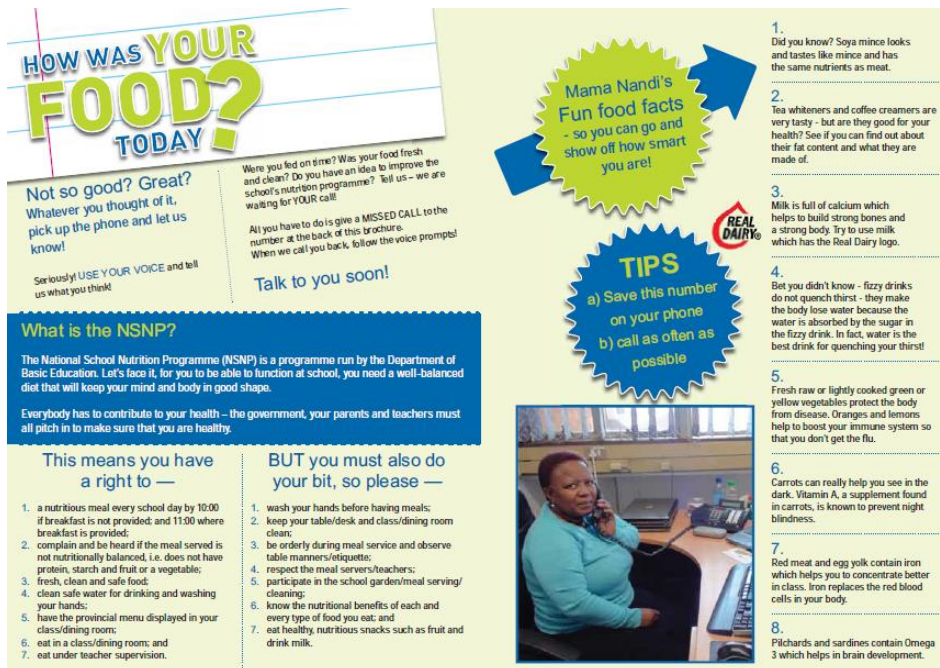


Figure 4: Revised Learner feedback service pamphlet

7.2.2 Coordinator reporting service

The NSNP coordinators (teachers) at schools use the Coordinator reporting service to give daily feedback on the meals served in their schools and provide details such as the menu of the food served, the time of feeding and number of learners fed. Since each province serves a different menu (a protein, a starch and a fruit/vegetable item) each province was consulted with respect to their menu options and the design was adapted for each province to ensure that the relevant options are posed in the IVR per province. In the event that the coordinator served a food item that is not available as an IVR option, a recording feature is provided to capture the food item served via an audio message. The coordinators may also leave general audio messages with regard to feeding-related issues at their school.

Since the call flow involves input on different types of data such as dates and numbers, specific error recovery checks were included in the IVR to cater for wrong inputs (i.e. checking for date formats and number of digits allowed). The school coordinators are also provided with the feature to call and report the prior days' statistics, if they were unable to provide a report on a particular day.

7.2.3 Web monitoring interface

The web monitoring interface allows the NSNP officials to not only obtain complete information on the feeding-related questions answered by the learners but also listen to the audio messages left by the learners and/or view the transcribed audio messages. The monitoring interface provides various filters to view the data based on provinces, schools, and time period (date).

The web monitoring interface also allows the NSNP officials to view the data provided by the school coordinators, to filter the data on school, province and time period level, and to download the data for analysis and/or reporting purposes.

7.3 Development

Development entailed the creation of two fully fledged IVR services and a web service. Firstly, an IVR service was developed for the NSNP school coordinators to report on the food they served to learners, and a web site was developed for DBE to access IVR-based coordinator feedback reports. An SMS reminder service was also created to remind coordinators to leave reports. Secondly, an IVR service was developed for learners to report on the food that they were served, and a web site was also developed for DBE to access IVR-based learner feedback reports.

Both IVR services were developed and provisioned using the existing Lwazi telephony platform. A generic content management system was created for use by both the web services, enabling information gathered during the IVR calls to be displayed and accessed. The content management system provides access to all user input captured during a call,

including touch-tone based feedback, and audio recordings. Lastly, a transcription workflow and interface was included in the content management system enabling all audio recordings to be transcribed.

A system monitoring web interface was also developed, providing a summarised view of calls received on both services, together with various indicators for call success.

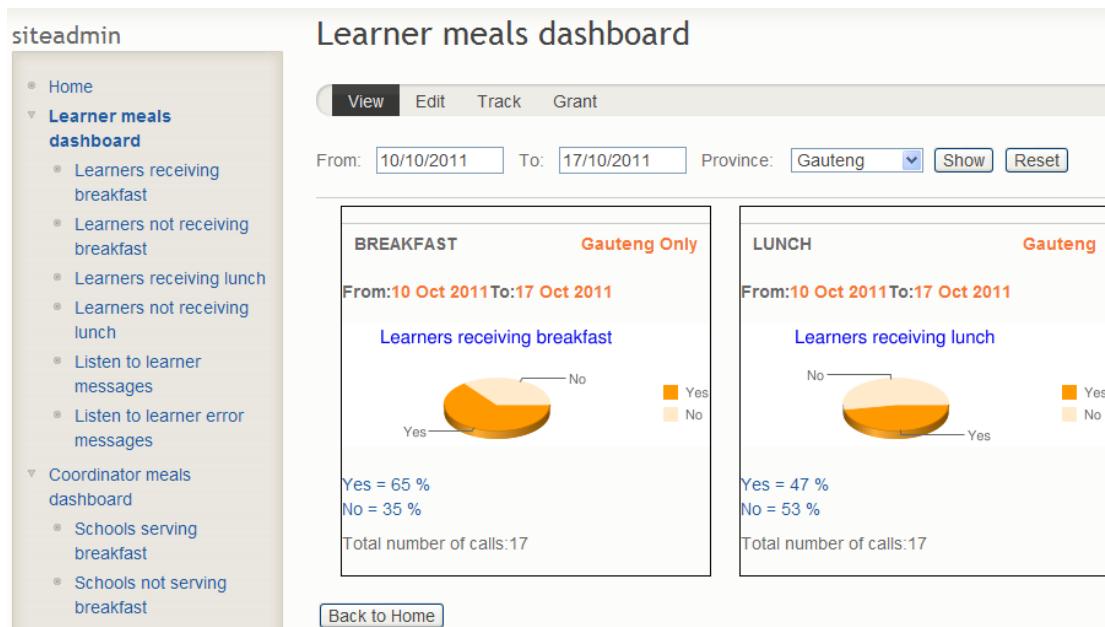


Figure 5: Web interface: Learner feedback service

In 2012, scalability enhancements were made to the call-back functionality of the Lwazi telephony platform, enabling multiple call-backs to be made simultaneously based on the number of available outgoing phone lines.

This new functionality facilitated the piloting process as multiple demonstrations of the system could take place simultaneously. Whereas previously, two calls could be made simultaneously to the system, the new functionality allowed up to six calls to be placed simultaneously.

7.4 Deployment

The DBE services were chosen for full deployment which took place in three phases:

The **phase 1** deployment took place in two schools in Gauteng, selected by the DBE based on a set of criteria supplied by the CSIR. The phase 1 Learner service was piloted in Sepedi, isiZulu and English.

The **phase 2** deployment took place in North West and Mpumalanga. Setswana and Afrikaans were added to the languages in which the Learner service was available for these pilots.

The **phase 3** deployment took place in Mpumalanga, Eastern Cape, Free State and Limpopo in the remaining official languages.

The deployment took place by means of *in situ* piloting of the Learner feedback and Coordinator reporting services. The main aims of the pilots were to –

- introduce the service to the learners and coordinators in the identified schools;
- demonstrate to the learners and coordinators how the service works; and
- encourage the learners and coordinators to use the systems by reporting on the meals they receive or serve at these schools.

The pilot sites were carefully selected to ensure a spread of languages. The actual piloting was preceded by visits to the pilot sites to build rapport with the principals and coordinators and determine the suitability of the sites.

The following criteria were used to select the pilot sites:

- The target audience was learners between 10 and 14 years of age (i.e. in Grades 8 to 10). Junior Secondary Schools were therefore preferred.
- Where possible, two schools (per province) which fall in the same district were selected – in order to facilitate logistical arrangements and obtaining the required permissions.
- The required NSNP infrastructure was required to be in place (i.e. NSNP coordinators, food handlers, good district participation) – this would ensure that the focus was on the services and not on NSNP implementation problems.
- The school needed to house a substantial number of learners – to increase the potential user base as much as possible.
- The area needed to have reasonable cell phone network coverage.

The details of the sites selected for piloting are summarised in Table 3.

Table 3: List of schools from all 3 phases

Province	Site (school name)	District	Coordinator's details	Languages	Date of pilot	IVR Tel No.
Phase 1						
Gauteng	Kokotla Junior Secondary School	Soshanguve	Ms Modise	English, Sepedi, isiZulu	27 Jul 2011	012 841 4952
Gauteng	Ntsako Junior Secondary School	Soshanguve	Ms Seloane	English, Sepedi, isiZulu	29 Jul 2011	012 841 4940
Phase 2						
North West	Inkonjane Middle School	Klipgat	Ms Maloka	English, Setswana, isiZulu, Afrikaans, Sepedi	2 Nov 2011	012 841 2163
North West	Thuto Pele High School	Letlhabile	Mr Mbedzi	English, Setswana, isiZulu, Afrikaans, Sepedi	3 Nov 2011	012 841 4802
Mpumalanga	Nqobangolwazi High School	Morgenzon	Ms Dlamini	English, Afrikaans, Setswana, isiZulu, Sepedi	7 Nov 2011	012 841 4572
Mpumalanga	Reggie Masuku High School	Ermelo	Mr Nhlambi	English, Afrikaans, Setswana, isiZulu, Sepedi	2 Feb 2012	012 841 4595
Phase 3						
Eastern Cape	Cowen Senior Secondary School	Port Elizabeth	Ms Mgaga	isiXhosa, English, isiZulu, Afrikaans, Sesotho	29 Aug 2012	012 841 2192

Province	Site (school name)	District	Coordinator's details	Languages	Date of pilot	IVR Tel No.
Eastern Cape	Khwezi Lomso High School	Port Elizabeth	Ms Nkumba	isiXhosa, English, isiZulu, Afrikaans, Sesotho	30 Aug 2012	012 841 2204
Free State	Iketsetseng Secondary School	Fezile Dabi	Ms Mavundla	Sesotho, English, Afrikaans, isiZulu, isiXhosa	7 Sep 2012	012 841 2214
Free State	Cedar Secondary School	Fezile Dabi	Mr Mthimkhulu	Sesotho, English, Afrikaans, isiZulu, isiXhosa	27 Aug 2012	012 841 2239
Limpopo	Tshiawelo Secondary School	Vhembe	Mr Neluvhola	Tshivenda, English, Sepedi, Xitsonga, Setswana	5 Sep 2012	012 841 2564
Limpopo	Marimane Secondary School	Vhembe	Mr Nwandule	Xitsonga, English, Sepedi, Tshivenda, isiZulu	4 Sep 2012	012 841 2701
Mpumalanga	Mabusa Besala Secondary School	Nkangala	Mr Rakwena	isiNdebele, isiZulu, English, Sepedi, Setswana	23 Aug 2012	012 841 2807
Mpumalanga	Mthombo Secondary School	Enhlanzeni	Ms Kekana	siSwati, English, isiNdebele, Xitsonga, isiZulu	16 Aug 2012	012 841 2816

7.4.1 Learner feedback service

7.4.1.1 Introduction

In order to gain entrance to the schools, the HLT team sent project introduction and description letters (an example is attached as annexure A) to the school principals.

In phases 1 and 2 the district offices were not involved in the process. This was rectified in phase 3 by sending the letters to the schools via the district office. The HLT team paid courtesy visits to many of the pilot schools to introduce the CSIR and the project, prior to actual piloting. During phases 1 and 2 piloting took place on one day. This was extended to two days in phase 3 depending on the proximity of the school from Pretoria. On the first day of the pilot visit, the district was provided with a verbal presentation of the project. District managers introduced the HLT team to the school principal, who in turn introduced them to the school coordinator. In phases 1 and 2, however, members of the national DBE NSNP unit accompanied the HLT team to the schools during the courtesy visits (and on piloting days) and introduced the team to the school principal.

7.4.1.2 Methodology

Piloting at the phase 1 schools entailed requesting random learners to test the system during their break times. However, this was a bit problematic as the environment was not controlled, leading to chaos. Following our experience in phase 1 pilots, piloting at the phase 2 schools entailed requesting two classrooms to be allocated with a number of learners who could test the system. We found this more organised and the piloting progressed more smoothly. Piloting at the phase 3 schools entailed selecting two learners per class to be Lwazi representatives “ambassadors”. The “ambassadors” were provided with a phone to test the system. This approach was taken so that the “ambassadors” would be comfortable enough with the system to demonstrate it to their fellow learners, and be able to answer questions that may arise when they introduced the system to their classmates.

7.4.1.3 Observations

The following represents a consolidation of the observations made across all three phases:

- Some learners asked if they could send “please call me” messages.
- Most of the learners asked if they could leave a message about their food preference. This was after they had called the service. The research team explained that there was a prompt which asked them to leave a message.
- Some learners were not sure if the service was free even though it had been explained that they needed to give the system a missed call.
- Some learners struggled to follow the prompts.
- Some learners had cell phones but they did not have airtime.

- Some learners complained that the “lady from the system” was asking them the same question over and over again.
- Many learners did not say the name of their school. When asked why not, they said they had not heard the prompt. Others said it was because they had assumed that the system knew the name of the school.
- Most learners were comfortable with the call-back concept and the DTMF input.
- Many learners finished the DTMF prompts but did not progress to the school name and food comment prompts.
- Some learners were concerned about the number that they should use to report the school meals, they were saying “but this number is not our number, how can we use it to report meals on the number that is not ours?”
- Learners were enthusiastic to test the application.
- Learners understood that the application was free, and all they needed to do was give it a missed call and it would call them back.
- Some learners were answering using their voice instead of using the key-pad.
- In Gauteng schools, some learners chose Sepedi, only 2 learners chose English. When asked why they chose English, one said it was because teachers use English to teach them and the other said she liked it and others said they chose it because they did not understand the other languages in the options provided. Learners who chose Sepedi said they chose it because it is their home language.
- Some learners were struggling to give the system a missed call and needed assistance.
- As is common with testing such services most learners responded to the system’s call by saying “Hello”.
- Some learners were initially shy to test the system and the teachers had to encourage them.
- Some learners did not want to test — when asked why, they said that they did not want other learners to hear what they were saying. In some schools it was especially the girls who seemed shy to test the system. Others said they would call the system when at home.
- Some learners who had mobile phones used these to call the system.
- The class got noisy at times, thus learners sometimes struggled to hear the system voice.
- A number of learners did leave a message on the meals at their schools when prompted to do so by the system.
- When asked whether they had enjoyed using the system, some learners were not sure if they had enjoyed it. They said it was okay, they could hear it.
- Upon being asked if anything was unclear, all learners responded that nothing was unclear.
- Some learners asked whether, by giving feedback, things would change in their school meals.
- When asked if they would recommend it to other learners, they said “yes”.

- When asked how frequently and when they would call the system, many learners said they would call after school using their own mobile phone or a family member's phone. Upon further probing some learners said that they might not be able to call every day.
- In one school learners were rushing to get out of the classrooms. When asked why, they said they were writing exams in the afternoon and wanted to study. This was despite the CSIR team having arranged with the school to pilot the services and being informed that it was in order to do so.
- Learners were given pamphlets so that they could distribute them to class mates.
- Some learners said the service was easy to use.
- Many learners seemed to like the system.
- To encourage confidentiality, learners had to be reminded not to leave their names.
- Learners wanted to discuss how unhappy they were with the food offered at the school.
- Learners wanted to know how soon they would see results when they started to use the service.
- Learners said the service was okay and that they would use it. They thought it was easy to follow the prompts.
- Some learners laughed every time other learners were leaving a message for Mama Nandi.



Figure 6: Learners at Iketsetseng Secondary School



Figure 7: Learners at Cedar Secondary School

7.4.1.4 Challenges

In addition to piloting in a class where learners were sitting, the system was piloted to other learners who were in groups outside the classroom. It was difficult to find one or two learners at a time. The noise levels were high and caused learners who were testing the system to struggle to hear the prompts.

When proceeding to a more random selection of learners, it was difficult for the piloting team (usually consisting of four persons) to manage all the learners who were willing to try the system (one person per phone dealing with many learners and at the same time monitoring the learner using the cell phone).

On a number of occasions, the system did not “call back” and the development team had to be contacted to check and rectify the problem. The technical team attributed this to the volume of calls due to the following:

- Although three numbers had been allocated for the piloting, the piloting team used only the number allocated to the school to pilot the system – this caused a flooding of the one line and resulted in delays in the system calling the users back.
- In planning the piloting day the technical team allocated three lines which could be comfortably accessed by six calls. When learners started testing the system using their own phones in addition to the six phones, on the one line, the line became overloaded and this resulted in delays in the system calling the users back. This impacted negatively on the users’ experience of the service.

When the system took long to call back, it resulted in learners losing interest in the service. The piloting team encouraged learners to keep calling the system. Some were patient in doing so while others showed no interest.

Another challenge was the language barrier. Some members of the research team could only communicate in English, which made some learners uncomfortable.

Some learners were studying for exams and indicated that was why they did not want to test the system. Others indicated they would test the system after writing exams. The timing of the pilot had a negative impact on potential future usage.

The learners grouped themselves into groups during the demonstration, and they were shouting at each other saying “I told you to press 1 and you just pressed 2”, and that affected other learners because they could not hear what was being said by Mama Nandi. Not everyone got the chance to experience the service and some were not happy about that.

Sometimes the class became chaotic and there was too much noise. The team and a teacher had to control them and/or some learners opted to test the system outside the classroom.

Some ambassadors did not want to test the service (phase 3). This was potentially problematic as they needed to demonstrate the system to other learners in their class. The researcher did find out if the co-ambassador from the same classes tested the service or not.

7.4.1.5 Opportunities

Peer education: one group (phase 3) was innovative and used the learners who had used the system to share their experiences with the rest of the class. The learners also motivated each other and stressed the seriousness of the service.

Availability of posters: each class was given a poster as a reminder for the learners to call the service daily.

Availability of teachers: Their presence helped control the learners and they were generally very attentive and took the pilots seriously.



Figure 8: Piloting at Mthombo Senior Secondary School

7.4.2 Coordinator reporting service

7.4.2.1 Introduction

Each school that was selected by the National NSNP office has a teacher designated to coordinate the running of the NSNP in that school. In a majority of these schools this teacher is also a Life Orientation instructor.

An application was built for school coordinators in order to facilitate their reporting on feeding provided in their schools. It was assumed that since the service would be paperless, it would be quick and easy for them to use. The NSNP school coordinator service requires the user (school coordinator) to give a missed call to the system. The system would call the user back. It would then ask a series of questions, requiring DTMF responses by the user.

All NSNP school coordinators were registered on the Lwazi system using their personal cell phone numbers. This enabled the service to recognise for which school they were reporting. A unique PIN was also allocated to each coordinator. Should the coordinator be unable to use the registered phone, s/he was at liberty to use any other available phone, and use his/her PIN as an identifier.

7.4.2.2 Methodology

The NSNP Coordinator service was piloted on the same day that the Learner service was piloted at a school. In most cases the coordinators met the piloting team in the principal's

office, and the principal was present when the service was demonstrated to the school coordinators. Each pilot began with an introduction of the CSIR and HLT. The same methodology was used at all schools for consistency sake. The introduction was followed by a full description of project Lwazi II.

A demonstration of the service was made after the introduction. The research team called the coordinator line and when the service called back, the phone was placed on speaker so that both the coordinator and the principal could hear. We also did this so that they could see how we responded to the prompts. After the demonstration using the project phone, the coordinator was given a PIN and was then asked to call the service using their cell phone which had been registered on the HLT service prior to the pilot.

7.4.2.3 Observation and comments

- The coordinators did not show any signs of struggling or lack of understanding. They seemed to follow the instructions with ease.
- The coordinators were happy with the system and promised to use it.
- The HLT team showed the coordinators the website and also how the district managers will access information about schools in their area.
- Some coordinators used cell phones that were not registered. This gave them an opportunity to learn how to use the PIN.
- The Coordinator reporting service was not piloted at Cedar Secondary school.

7.4.3 Involvement of district NSNP officials

7.4.3.1 Introduction

In phase 3, district-level NSNP officials were also included in the piloting of these services. Five districts were involved, namely: Vhembe in Limpopo, Port Elizabeth in Eastern Cape, Ehlanzeni and Nkangala in Mpumalanga and Fezile Dabe in Free State. The districts were selected by the national NSNP office according to the research team's request that the system be piloted in all official languages.

7.4.3.2 Methodology

Each district was initially contacted via telephone. A letter stating the details of the project and the intentions to visit the school was sent (attached as annexure A). An initial visit to the district office was done prior to piloting. During these visits, district managers were given details on the project background, pilots and future possibilities. Engagement with the district officials took place on the pilot day. The pilot dates were circulated beforehand and confirmed a week before each pilot.

On arriving in the province for piloting, the research team visited the district offices before proceeding to the school. District managers and colleagues were present at each district

office. The pilot started with an explanation of the project background to the colleagues. This was followed by a demonstration of the Learner feedback service and an explanation of the Coordinator reporting service. This was done in order to demonstrate that calls were reflected almost immediately on the web monitoring interface which the district coordinators would use to track reports from schools in their district.

An initial demonstration of the web interface was done. All the web pages were explained, and the test calls that had just been made were shown to those present. They were also shown how to retrieve information for inclusion in reports. Managers who had internet access were asked to log onto the system using their own laptops and the logon credentials which had been supplied to them.

Even though four out of five district managers had internet access, they indicated that it is not always available. At times, they use their personal internet connections which will not motivate them to use the service.

7.4.3.3 Questions

The following questions were raised by the district coordinators:

- Could they access the service using any machine, or had the system been installed on their office machine where we had piloted?
- Were they allowed to show the system to other colleagues in the district office?
- Would the service replace any of the existing national forms?

There were also a number of questions about the different functions and fields on the web interface.

7.4.4 SMS Report reminder service

The SMS report reminder service was piloted within the NSNP unit at national level in January 2011, and within the Chief Directorate to which the NSNP unit belongs, in September 2011.

The service was demonstrated to the target users on these occasions and several follow-up calls were made to ensure that the users understood the functionality of the system and how to use it.

The system is at present still in use at the national NSNP offices and can be used to disseminate information on upcoming events and send out reminders for reporting deadlines.

7.4.5 Conclusion

The piloting of the Lwazi II DBE NSNP services took place over a 24-month period (January 2011 – December 2012). The NSNP Lwazi services were introduced to potential users by means of demonstrations and users were encouraged in various ways to use the services.

Feedback on the usability and user experience of the service will be provided in the evaluation sections which follow.

7.5 Call log analysis

In the following section we present an in-depth call log analysis of the calls placed to the Learner feedback service and the Coordinator reporting service. Call log analysis presents a quantitative view towards understanding the usage of a service and is complementary to the qualitative aspect of analysing overall user experience through focus groups and interviews. Various metrics such as volumes of calls, call durations, call patterns such as final exit states, invalid entries and timeouts that occur in a call, are analysed to obtain a thorough understanding of the interaction a user has with the system during a call.

7.5.1 Learner feedback service

The call log analysis showcased in this section covers the calls made to the service by learners in 14 schools from May 2011 to December 2012⁵. Learner feedback service phase 1 was piloted from July to October 2011. In phase 2 (November 2011 – July 2012) the service was introduced to four more schools in Mpumalanga (two schools) and North West (two schools), and focus groups were also being conducted during this period. In phase 3, eight more schools were added during the period August to November 2012; these schools were in Mpumalanga (two more schools), Eastern Cape (two schools), Free State (two schools) and Limpopo (two schools). In reflecting on these results, it is important to note that the running period of the service at each of these 14 schools differs. In Gauteng the service ran for one year and five months, in North West and one school in Mpumalanga (Nqobangolwazi) for a year, whereas in the schools piloted in phase 3 it ran for five months. For each school the analysis was conducted only on calls made by the learners after the piloting. This was to ensure that the calls represented in this analysis were not calls made during demonstration of the service to the learners (service being introduced to learners during piloting). Table 3 shows the dates on which the service was piloted in each school per province.

⁵ Calls were analysed up to 7 December 2012, which was the last day of the 2012 school year.

7.5.1.1 Call volume

The total number of calls made by learners between July 2011 and December 2012, was **3 081 calls**. Figure 9 below shows the time distribution of the calls received by the service during this period.

We noticed that most calls were made during the months when the service was introduced to schools, for phase 1 the service was piloted in July to October 2011. For phase 2, four schools were piloted between November 2011 and July 2012 and for phase 3, eight more schools were piloted between August and November 2012.

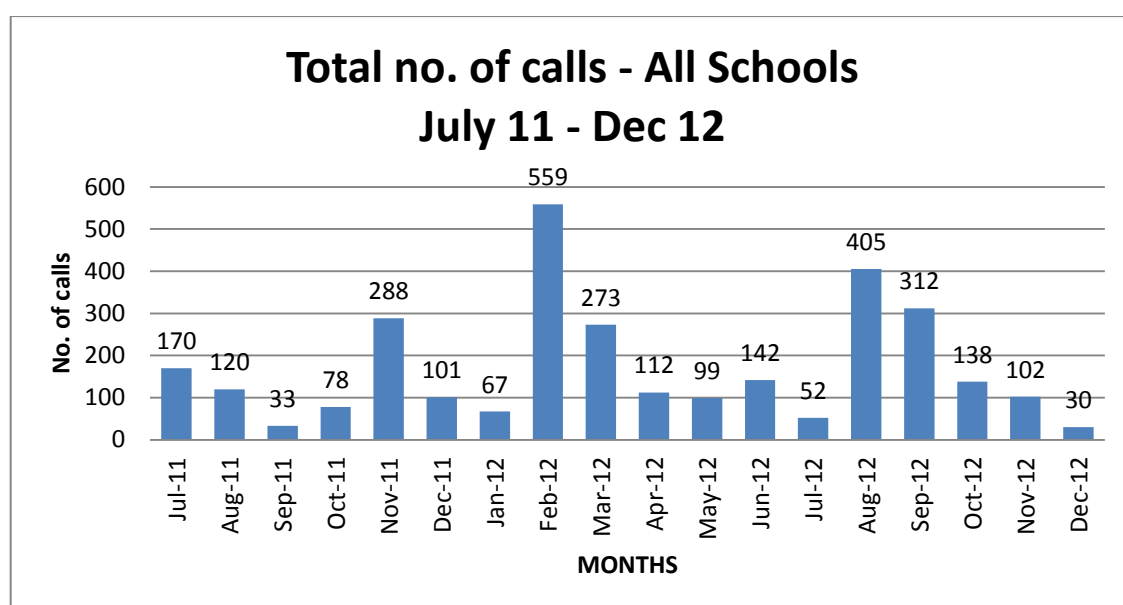


Figure 9: Total call distribution per month

Table 4 below illustrates the spread of calls across provinces. It can be observed that the most calls came from learners in the provinces where the service was introduced a year ago. These seem to taper off, however, between September and December 2012.

Table 4: Total call distribution per month per province

Month	Gauteng	Eastern Cape	Free State	Mpumalanga	Limpopo	North West
Jul-11	170	-	-	-	-	-
Aug-11	120	-	-	-	-	-
Sep-11	33	-	-	-	-	-
Oct-11	78	-	-	-	-	-

Month	Gauteng	Eastern Cape	Free State	Mpumalanga	Limpopo	North West
Nov-11	33	-	-	35	-	220
Dec-11	26	-	-	46	-	29
Jan-12	8	-	-	28	-	31
Feb-12	101	-	-	323	-	135
Mar-12	108	-	-	78	-	87
Apr-12	34	-	-	47	-	31
May-12	41	-	-	41	-	17
Jun-12	50	-	-	68	-	24
Jul-12	26	-	-	20	-	6
Aug-12	49	46	23	108	150	29
Sep-12	0	39	59	67	126	21
Oct-12	22	35	8	51	7	15
Nov-12	13	8	3	29	34	15
Dec-12	0	3	0	17	3	7
Total	912	131	93	958	320	667
%	29.60 %	4.25 %	3.02 %	31.09 %	10.39 %	21.65 %

The distribution of the total calls on a per school basis is illustrated in Table 5 below. From this again we observe that the schools with the highest number of calls are those piloted a year ago. These are Kokotla (15.06%) and Ntsako (14.54%) (both from Gauteng), followed by Nqobangolwazi (13.60%) from Mpumalanga and Thuto Pele (12.63%) from North West.

Table 5: Total call distribution per month per school

Month	School													
	Gauteng		North West		Mpumalanga				Eastern Cape		Free State		Limpopo	
	Kokotla	Ntsako	Inkonjane	Thuto ⁶	Nqoba ⁷	Reggie ⁸	Mthombo	Mabusa ⁹	Cowen	Kwezi ¹⁰	Iketseng	Cedar	Marimane	Tshiwela
Jul-11	90	80	-	-	-	-	-	-	-	-	-	-	-	-
Aug-11	57	63	-	-	-	-	-	-	-	-	-	-	-	-
Sep-11	13	20	-	-	-	-	-	-	-	-	-	-	-	-
Oct-11	46	32	-	-	-	-	-	-	-	-	-	-	-	-
Nov-11	27	6	77	143	35	-	-	-	-	-	-	-	-	-
Dec-11	19	7	2	27	46	-	-	-	-	-	-	-	-	-
Jan-12	4	4	2	29	28	-	-	-	-	-	-	-	-	-

⁶ Henceforth we refer to the Thuto Pele secondary school as “Thuto” in the tables and figures.

⁷ “Nqoba” refers to Nqobangolwazi Secondary School.

⁸ “Reggie” refers to Reggie Masuku Secondary School.

⁹ “Mabusa” refers to Mabusa Besala Secondary School.

¹⁰ “Kwezi” refers to Kwezilomso Secondary School.

Month	School													
	Gauteng		North West		Mpumalanga				Eastern Cape		Free State		Limpopo	
	Kokotla	Ntsako	Inkonjane	Thuto ⁶	Nqoba ⁷	Reggie ⁸	Mthombo	Mabusa ⁹	Cowen	Kwezi ¹⁰	Iketseng	Cedar	Marimane	Tshiwela
Feb-12	62	39	64	71	116	207	-	-	-	-	-	-	-	-
Mar-12	43	65	33	54	28	50	-	-	-	-	-	-	-	-
Apr-12	26	8	13	18	29	18	-	-	-	-	-	-	-	-
May-12	34	7	4	13	12	29	-	-	-	-	-	-	-	-
Jun-12	21	29	15	9	52	16	-	-	-	-	-	-	-	-
Jul-12	13	13	5	1	17	3	-	-	-	-	-	-	-	-
Aug-12	3	46	22	7	18	0	62	28	20	26	-	0	-	-
Sep-12	0	0	12	9	8	1	44	14	11	28	30	23	82	150
Oct-12	1	21	12	3	21	1	21	8	5	30	3	29	7	44
Nov-12	5	8	13	2	8	7	12	2	2	6		5	6	28
Dec-12	0	0	4	3	1	9	7	0	1	2		3	0	3
Total	464	448	278	389	419	341	146	52	39	92	33	60	95	225

Month	School													
	Gauteng		North West		Mpumalanga				Eastern Cape		Free State		Limpopo	
	Kokotla	Ntsako	Inkonjane	Thuto ⁶	Nqoba ⁷	Reggie ⁸	Mthombo	Mabusa ⁹	Cowen	Kwezi ¹⁰	Iketseng	Cedar	Marimane	Tshiawela
%	15.06 %	14.54%	9.02%	12.63%	13.60%	11.07%	4.74%	1.69%	1.27%	2.99%	1.07%	1.95%	3.08%	7.30%

In Table 6 we further analyse this data to calculate the average calls per week which proportions the total calls received at a school with respect to the number of weeks the pilot has been running at the school. Noticeably, Tshiawela has the highest (16.8) number of calls/week, which may possibly be attributed to the fact that the piloting took place in September 2012 as compared to other schools in which the service was piloted in 2011 and early 2012.

Table 6: Average calls per week

	School													
	Gauteng		North West		Mpumalanga				Eastern Cape		Free State		Limpopo	
	Kokotla	Ntsako	Inkonjane	Thuto	Nqoba	Reggie	Mthombo	Mabusa	Cowen	Kwezi	Iketseng	Cedar	Marimane	Tshiawela
Calls	464	448	278	389	419	341	146	52	39	92	33	60	95	225
No. of pilot weeks	71.3	71.0	57.3	57.1	56.6	42.3	16.1	15.1	14.3	14.1	13.0	14.6	13.3	13.4
Calls/week	6.5	6.3	4.9	6.8	7.4	8.1	9.1	3.4	2.7	6.5	2.5	4.1	7.1	16.8

7.5.1.2 Call duration

Average call duration was found to be 1 min and 28 s. Figure 10 illustrates the distribution of call duration across all the schools. We observe that 35.8% of the calls end between 15 s – 1 min, with significantly 59% of the calls received by the service being of the duration 1 min 15 s – 3 min, and 4.9% of the calls ended with a duration greater than 3 min.

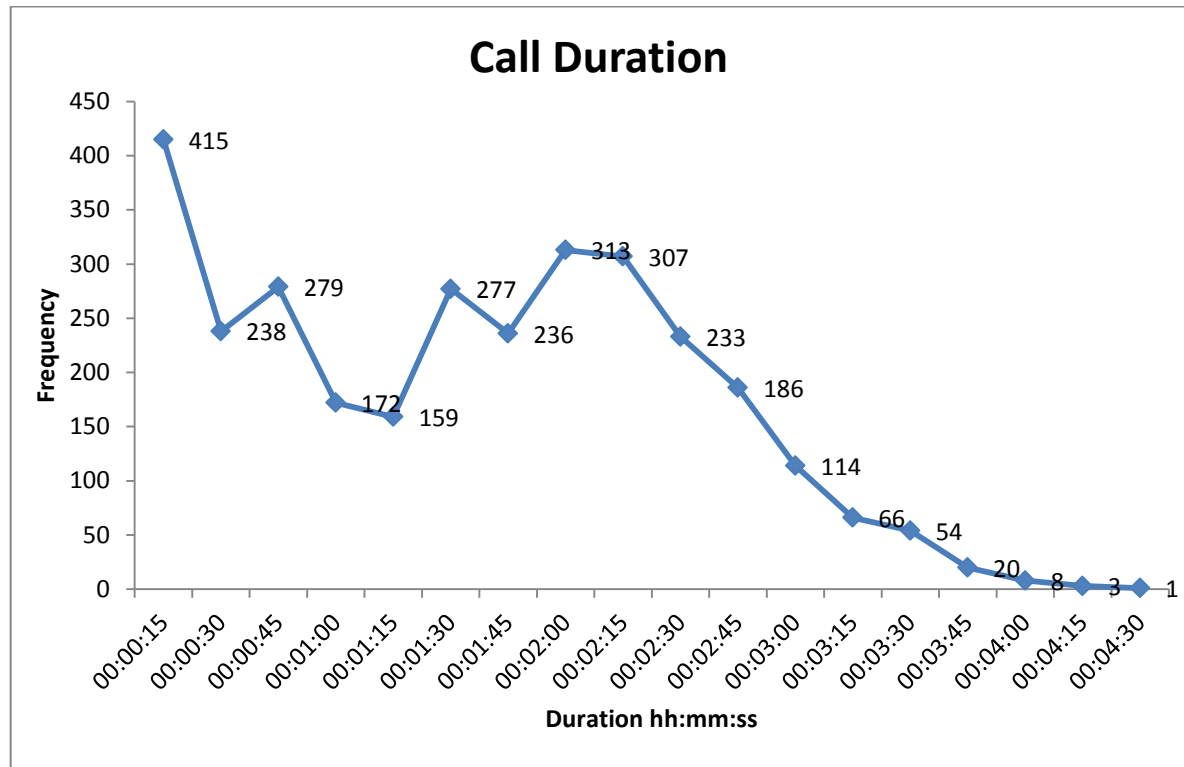


Figure 10: Call duration distribution

7.5.1.3 Usage frequency analysis

Across the period of July 2011 to December 2012, the service received a total number of 3081 calls which we further break down to perform a usage frequency analysis which shows the frequency of how often users call the service (repeat vs. one-time usage).

Table 7 illustrates this frequency analysis, where we have the “no. of calls per user”, which indicates the number of times a user has called the service, the “no. of unique callers” which refers to the number of unique callers¹¹ who called the service a particular number of times (e.g. 63 users who called the service between 6-10 times), “no. of calls” and “% of total calls” which presents the number of calls attributed to that category of usage and the proportion of the total calls it represents.

¹¹ Unique caller refers to the caller using a unique phone number, and not a particular individual per se who may have used multiple phone numbers to call the service.

We find that 466 calls were made by 466 unique callers (i.e. people who called the service once). Another 297 users called the service 2-3 times, and 70 users called the service 4-5 times over this analysis period. Strikingly, we find that approximately 10% of the calls were made by seven unique users who called 31-73 times and another 18% of calls can be attributed to users who called the service 11-30 times.

Based on caller identification we also find that 154 calls (4.9%) were made from a landline, 2632 calls (85%) were calls made using cell phones.

Table 7: Usage frequency analysis

No. of calls per user	No. of unique callers	No. of calls	% of total calls
1	466	466	15%
2-3	297	689	22%
4-5	70	302	10%
6-10	63	442	14%
11-20	28	399	13%
21-30	7	165	5%
31-40	3	96	3%
41-50	1	49	2%
51-60	2	105	3%
61-75	1	73	2%
Caller ID not identified ¹²	-	295	10%
TOTAL	938	3081	100%

7.5.1.4 Language choice

During a call the service provides learners with a choice of 4-6 languages upfront, with a “language menu”. These 4-6 languages chosen for the language menus vary across the provinces, taking into account the provincial language preferences as well as the local prevailing language preferences within a school and/or district.

¹² Caller ID logging experienced a technical error during the July-Sep 2011 period, which led to unreliable caller ID data for that period.

Across the six provinces, all 11 official languages are covered by the service, viz., English (eng), Sepedi (sep), Setswana (set), isiZulu (zul), Afrikaans (afr), Sotho (sot), siSwati (swa), Xitsonga (tso), Tshivenda (ven), isiXhosa (xho), and isiNdebele (nde). Learners choose a language of their choice from the language menu, and thereafter the interaction of the service with the learner continues in the chosen language. Table 8 shows a breakdown of the language choice made by the learners across all schools, and Figure 11 shows language choice in all schools in all provinces. We observe that learners do choose their home language, with African languages collectively being chosen in 56%, and English being chosen in 44% of the total calls.

Table 8: Language choice distribution per school per province

Lang	School														Total	%
	Gauteng		North West		Mpumalanga				Eastern Cape		Free State		Limpopo			
	Ntsako	Kokotla	Inkonjane	Thuto Pele	Mthombo	Nqoba	Reggie	Mabusa	Cowen	Kwezi	Iketseng	Cedar	Tshiwela	Marimane		
sep	113	131	17	37	-	13	5	1	-	-	-	-	10	-	327	10.61 %
afr	7	2	5	4	-	19	2	-	-	7	-	-	2	-	48	1.56%
eng	277	244	125	180	88	133	134	14	13	35	15	22	72	19	1371	44.50%
nde	-	-	-	-	2	-	-	35	-	-	-	-	-	-	37	1.20%
set	21	20	113	151	-	16	3	-	-	-	-	-	-	-	324	10.52%
sot	-	-	-	-	-	-	-	-	1	3	15	36	-	-	55	1.79%
swa	-	-	-	-	45	-	-	-	-	-	-	-	-	-	45	1.46%
tso	-	-	-	-	3	-	-	-	-	-	-	-	12	75	90	2.92%
ven	-	-	-	-	-	-	-	-	-	-	-	-	129	1	130	4.22%
xho	-	-	-	-	-	-	-	-	25	46	1	1	-	-	73	2.37%
zul	30	67	18	17	8	238	197	2	-	1	2	1	-	-	581	18.86%
Grand Total	3081															

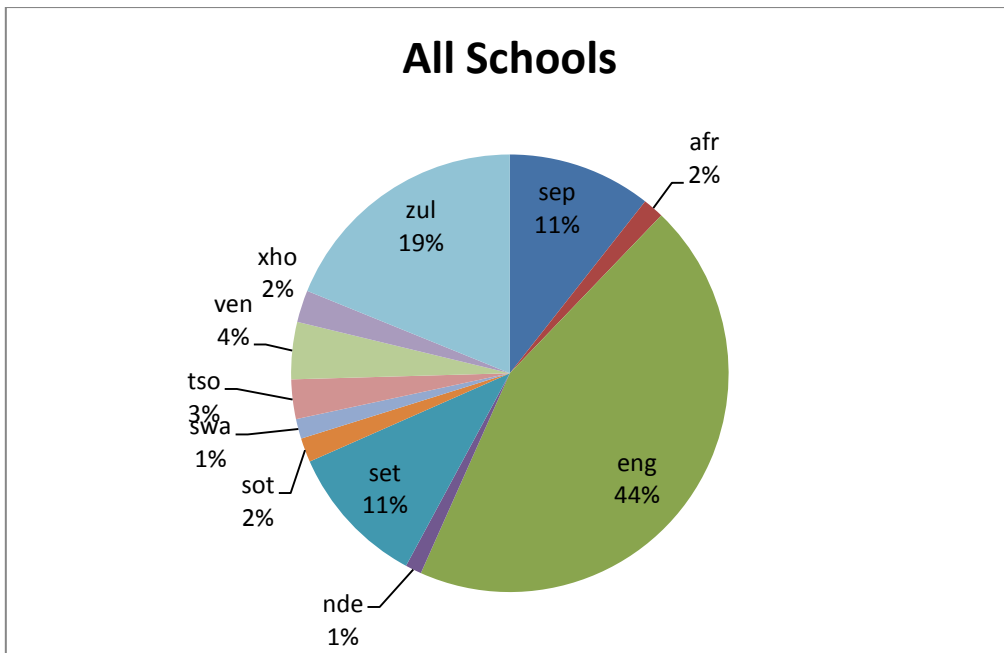


Figure 11: Language distribution in all schools and provinces

In examining the language choice we further analysed the school location in terms of the province, since each province has dominant languages, and often even within provinces there are specific areas where a particular language may be used more often than others. Figure 12 illustrates the language choice in terms of percentage; for Gauteng we find that English (57%) was by far the most chosen language followed by Sepedi (27%).

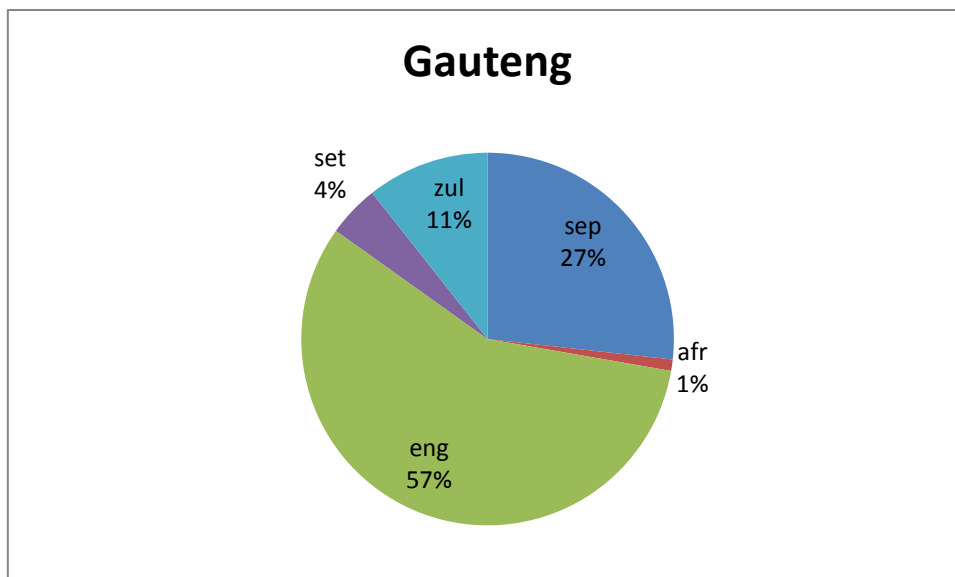


Figure 12: Language choice – Gauteng

Table 9 shows a breakdown of the language choice per school in Gauteng.

Table 9: Gauteng – Language choice distribution per school

Languages	Schools		Grand Total
	Kokotla	Ntsako	
sep	131	113	244
afr	2	7	9
eng	244	277	521
set	20	21	41
zul	67	30	97
Grand Total	464	448	912

Figure 13 below illustrates the language choice for North West where we observe that 46% of the learners chose English, and 53% chose African languages (with 40% choosing Setswana).

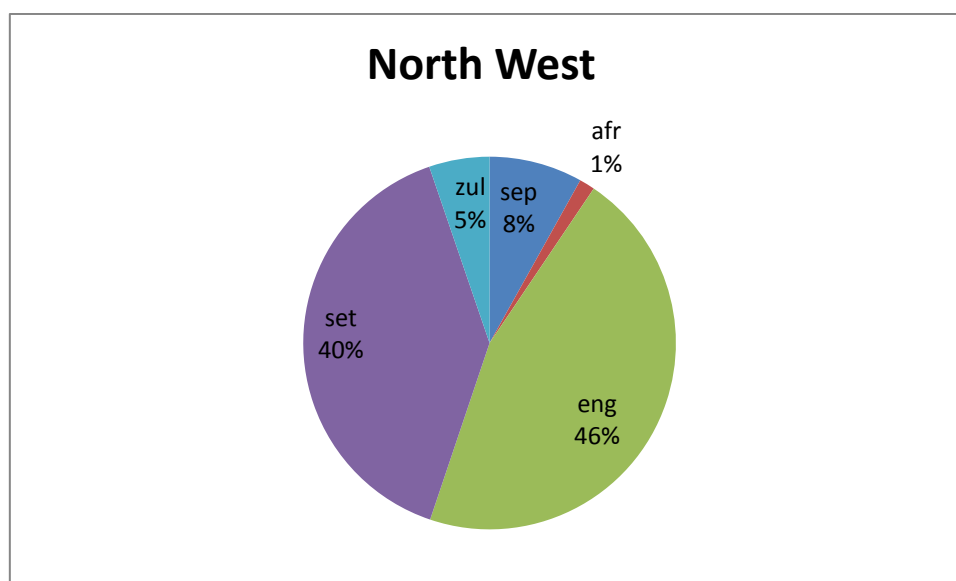


Figure 13: Language choice – North West

Table 10 shows a breakdown of the language choice per school in North West.

Table 10: North West – Language choice distribution per school

Languages	Schools		Grand Total
	Inkonjane	Thuto Pele	
sep	17	37	54
afr	5	4	9
eng	125	180	305
set	113	151	264
zul	18	17	35
Grand Total	278	389	667

Figure 14 illustrates the language choice for Mpumalanga. We notice that most learners prefer isiZulu (46%), followed by English (39%), with Sepedi, Afrikaans, isiNdebele, Setswana, siSwati and Xitsonga making up the remaining 15% of the calls.

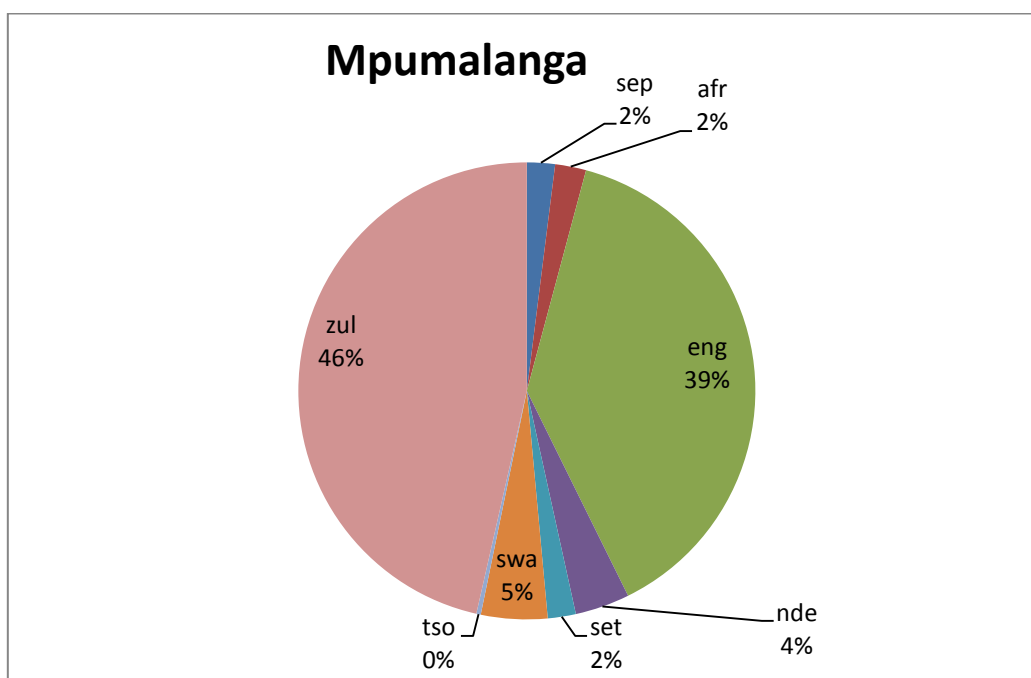


Figure 14: Language choice – Mpumalanga

Table 11 shows a breakdown of the language choice per school in Mpumalanga.

Table 11: Mpumalanga – Language choice distribution per school

Languages	Schools				Grand Total
	Mabusa Besala	Mthombo	Nqobangolwazi	Reggie Masuku	
sep	1		13	5	19
afr			19	2	21
eng	14	88	133	134	369
nde	35	2			37
set			16	3	19
swa		45			45
tso		3			3
zul	2	8	238	197	445
Grand Total	52	146	419	341	958

Figure 15 illustrates the language choice for Eastern Cape where we find that isiXhosa (54%) was by far the most preferred language, followed by English at 37%.

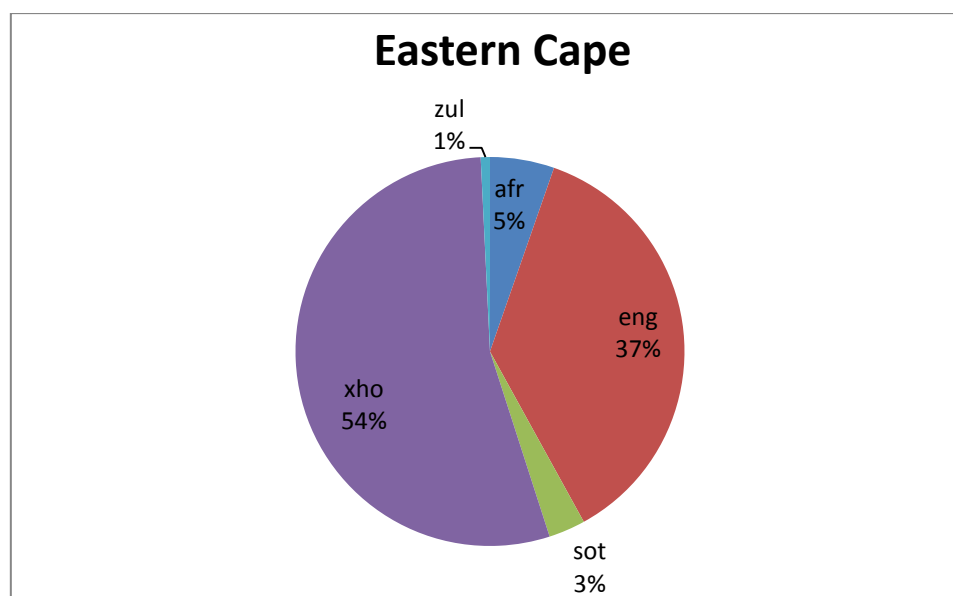


Figure 15: Language choice – Eastern Cape

Table 12 shows a breakdown of the language choice per school in Eastern Cape.

Table 12: Eastern Cape – Language choice distribution per school

Languages	Schools		Grand Total
	Cowen	Kwezilomso	
afr		7	7
eng	13	35	48
sot	1	3	4
xho	25	46	71
zul		1	1
Grand Total	39	92	131

Figure 16 illustrates the language choice for Free State where we find that Sesotho (55%) was by far the most preferred language, followed by English at 40%.

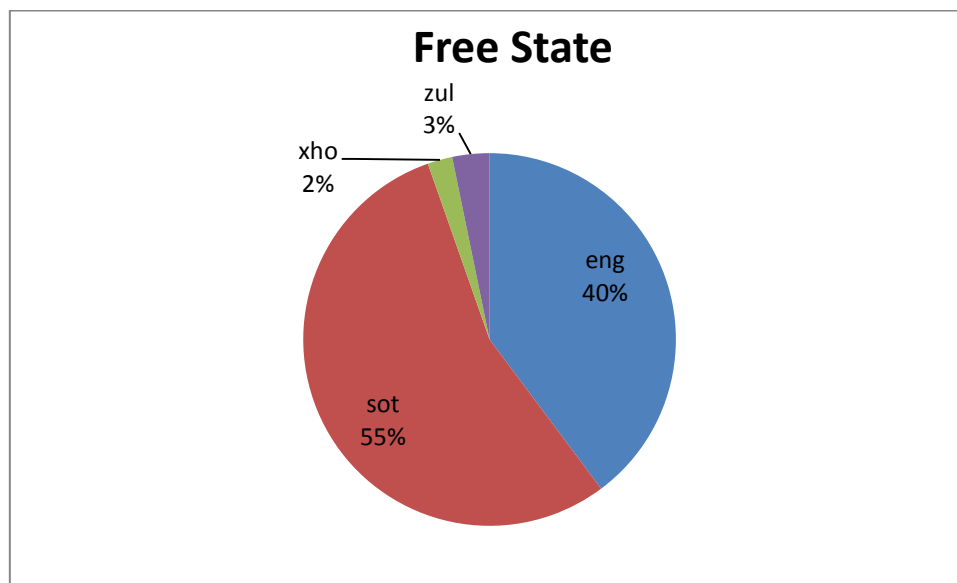


Figure 16: Language choice – Free State

Table 13 shows a breakdown of the language choice per school in Free State.

Table 13: Free State – Language choice distribution per school

Languages	Schools		Grand Total
	Cedar	Ikeketseng	
eng	22	15	37
sot	36	15	51
xho	1	1	2
zul	1	2	3
Grand Total	60	33	93

Figure 17 illustrates the language choice for Limpopo where we observed that Tshivenda (41%) was the most preferred language, followed by English (28%) and Xitsonga (27%).

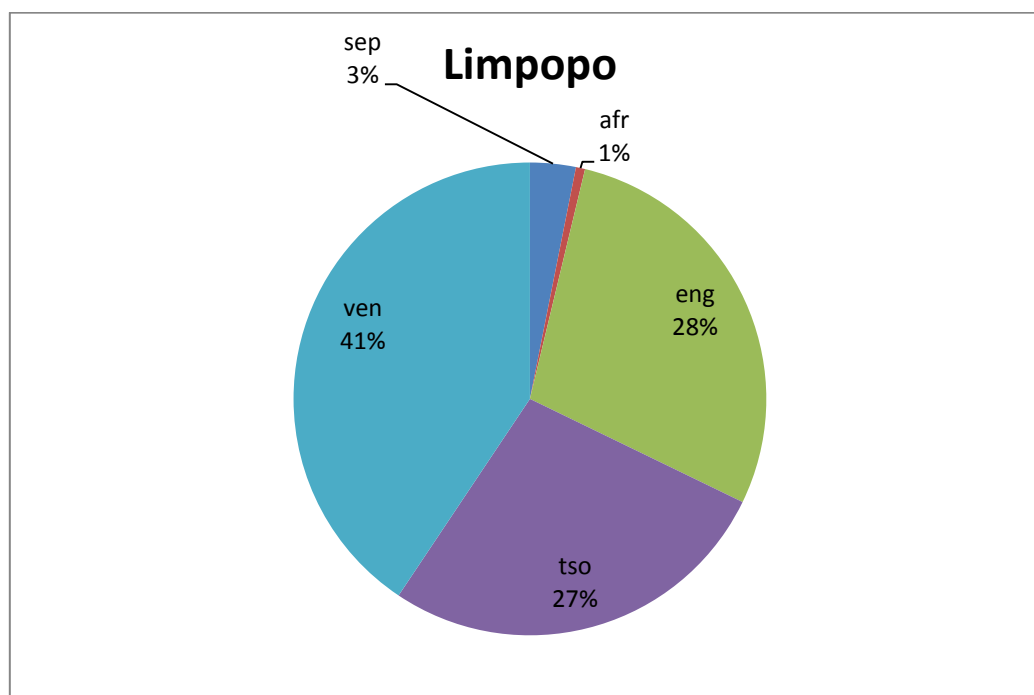


Figure 17: Language choice – Limpopo

Table 14 shows a breakdown of the language choice per school in Limpopo and Table 15 illustrates language choice per province across all provinces.

Table 14: Limpopo – Language choice distribution per school

Languages	Schools		Grand Total
	Marimane	Tshiawela	
sep		10	10
afr		2	2
eng	19	72	91
tso	75	12	87
ven	1	129	130
Grand Total	95	225	320

In Table 15 we provide an overall snapshot of the language choice across all provinces, which highlight the most preferred language per province.

Table 15: Language choice distribution per province

Languages	Provinces					
	Gauteng	North West	Mpumalanga	Eastern Cape	Free State	Limpopo
sep	27%	8%	2%	-	-	3%
afr	1%	1%	2%	5%	-	1%
eng	57%	46%	39%	37%	40%	28%
sot	-	-	-	3%	55%	-
xho	-	-	-	54%	2%	-
zul	11%	5%	46%	1%	3%	-
nde	-	-	4%	-	-	-
set	4%	40%	2%	-	-	-
swa	-	-	5%	-	-	-
tso	-	-	0%	-	-	27%
ven	-	-	-	-	-	41%

7.5.1.5 Final call state

The final call state represents the last dialog state (or step) in the call flow of the service where the user ended the call. Overall this final state gives an indication whether the learners were able to complete the task of providing feedback about their meals successfully. It also provides more insight into the possible areas of the call flow where users are dropping off, which might require further attention. Figure 18 illustrates the final call state for calls received across all the schools. We observe that out of 3 081 calls, 1 503 (48%) of those calls ended successfully with learners hanging up at the ‘goodbye’ state (0.18 Goodbye) where learners were thanked for their feedback. Some learners chose to hang up at the ‘0.2 Language Menu’ (18%) state where they were required to choose a language of interaction with the service.

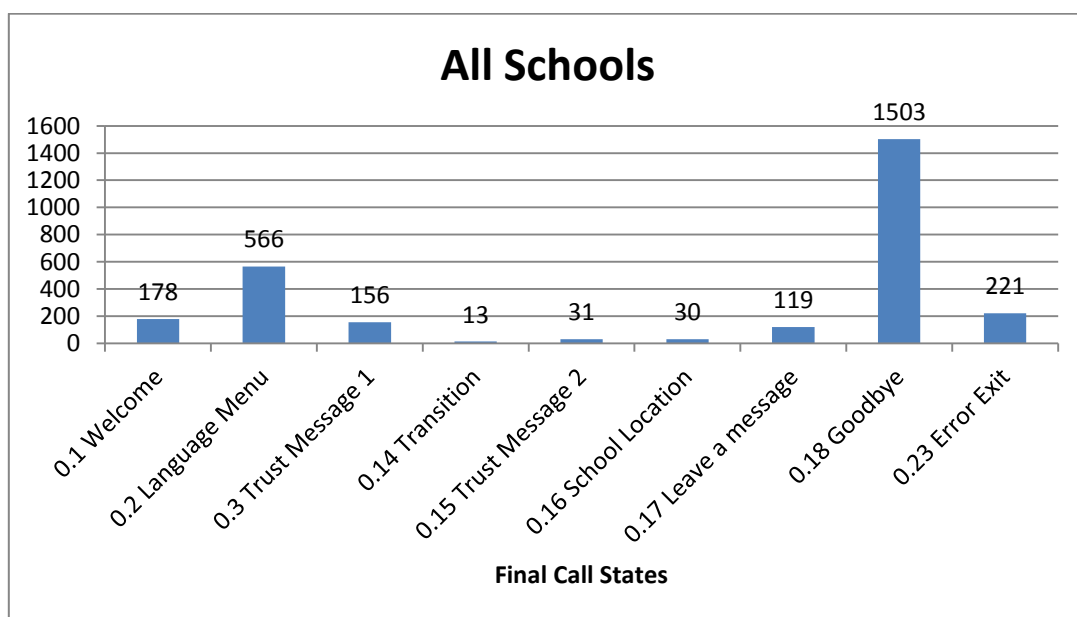


Figure 18: Final call state for all schools

In terms of proportion (shown in Table 16), we find that for 55% of the calls, the final state was a dialog state where learners had successfully completed the questions related to the school meals’ feedback (states 0.14-0.18).

We also find that 29% of the calls were dropped at the first three dialog states of the service (states 0.1-0.3), with the majority of these being dropped at the “0.2 Language Menu” state. There may be several reasons for this, such as that the learners were just trying out the system, or were not fully sure what to do and thus hung up as soon as they were asked to provide first input to the service (at state “0.2 Language Menu” the learner is asked to make a language choice by pressing a button).

Table 16: Distribution of final call states

Final States	No. of calls	%
0.1 Welcome	178	29%
0.2 Language Menu	566	
0.3 Trust Message 1	156	
0.14 Transition	13	55%
0.15 Trust Message 2	31	
0.16 School Location	30	
0.17 Leave a message	119	
0.18 Goodbye	1503	
0.23 Error Exit	221	
Other	264	9%

Another 7% of the calls ('0.23 Error Exit' state) are attributed to instances where learners did not make any input to the service more than 2 times and the call was thus ended by the service with appropriate feedback (see next section for further details). Examples of error exits include not pressing a button, or giving wrong input when prompted by the service (e.g. pressing a button that was not allowed at a particular menu option e.g. pressed 5 when the options were to press 1 or 2). The "Other" category of 9% refers to calls where the learners dropped the call in one of the dialog states (0.4-0.13) where questions related to the meals were still being asked, e.g. meal availability, tasty, on time, etc.

Table 17 shows a breakdown of the final states in the call flow of the service where the user ended the call across all schools.

Table 17: Final state distribution per school

Final Call States	School														Total	%
	Gauteng		North West		Mpumalanga				Eastern Cape		Free State		Limpopo			
	Ntsako	Kokotla	Inkonjane	Thuto Pele	Mthombo	Nqoba	Reggie	Mabusa	Cowen	Kwezi	Iketseng	Cedar	Tshiwela	Marimane		
0.1 Welcome	37	40	29	9	10	17	16	2		1	2	3	9	3	178	6%
0.2 Language Menu	95	65	51	57	51	70	59	11	6	25	12	9	39	16	566	18%
0.3 Trust Message 1	25	26	9	23	12	29	12	2	3	3	1		10	1	156	5%
0.14 Transition	3	2	1		1	2	1	1	0	0	0	1	0	1	13	0%
0.15 Trust Message 2	1	2		3	1	19	1	1	1	1	0	1	0	0	31	1%
0.16 School Location	3	3	2	5	1	8	5	0	0	1	0		2	0	30	1%
0.17 Leave a message	11	14	3	19	4	21	16	3	2	3	6	6	9	2	119	4%
0.18 Goodbye	172	203	154	219	40	195	192	30	20	51	9	32	121	65	1503	49%
0.23 Error Exit	62	43	13	24	13	24	19	2	1	1		1	18		221	7%
Other (Dialog states 0.4 -0.13)	39	66	16	30	13	34	20	0	6	6	3	7	17	7	264	9%
Grand Total	3081															100%

7.5.1.6 Invalid entries

An invalid entry (or wrong input) refers to a case where users press an invalid key when prompted by the service to choose from a defined set of keys at a particular dialog state. For example, many questions in the service provide the learner with the option to press 1 or 2; if a learner presses any other key besides 1 or 2 at such a question, this is logged as an invalid entry. For the first and second time when a learner provides an invalid entry, they are re-prompted by the service appropriately to answer the question with valid entries only. The third consecutive time, the service requests the learner to leave a message should they require further help in using the service and exits the call (this is termed as an ‘error exit’ represented by the “0.23 Error Exit” state).

Table 18 provides an overview of the overall invalid entries experienced across the service. We find that there were 90 calls (2.9% of total calls) which had invalid entries. We also find that of the 90 calls, in 68 calls (2%) the learners only made 1 invalid entry in the call, which implies that upon the first re-prompt, the learner successfully chose a valid entry.

Table 18: Invalid entries – All schools

Provinces	Schools	No. of invalid entries/call			Total
		1	2	3	
Gauteng	Kokotla	9	2	2	13
	Ntsako	10	2	0	12
North West	Inkonjane	7	1	1	9
	Thuto Pele	3	3	2	8
Mpumalanga	Mabusa Besala	1	0	0	1
	Mthombo	6	1	0	7
	Nqobangolwazi	16	3	0	19
	Reggie Masuku	6		1	7
Eastern Cape	Cowen	0	1	0	1
	Kwezilomso	3	1	1	5
Free State	Cedar Secondary	2	0	0	2
	Iketseng	1	0	0	1
Limpopo	Marimane	0	0	0	0

Provinces	Schools	No. of invalid entries/call			Total
		1	2	3	
	Tshiawela	4	1	0	5
Grand Total		68	17	7	90

7.5.1.7 Timeouts

A timeout refers to the case where a learner provides no input (does not press any key) when the service asks a question. Similar to the invalid entries, the service will re-prompt the learner twice for input using appropriately designed prompts, and at the third consecutive time the service will ask them to leave a message and exit the call ('error exit'). Table 19 provides an overview of the number of calls in relation to the frequency of timeouts. We find that 448 calls (14% of total) had more than 1 timeout, i.e. in 448 calls (15%) calls the learner did not provide any input at least once, while 387 calls (12%) had 2 timeouts and 38 calls (1%) had 3–6 timeouts. This provides an indication that although learners did timeout, after being re-prompted twice they were able to proceed further with the call by providing some kind of input.

Table 19: Timeouts – All schools

Provinces	Schools							Grand Total
		1	2	3	4	5	6	
Cedar		7	8	1	0	0	0	16
Cowen		9	3	1	0	0	0	13
Iketseng		4	7	0	0	0	0	11
Inkonjane		25	29	2	0	0	0	56
Kokotla		64	61	4	4	2	0	135
Kwezilomso		19	10	0	0	0	0	29
Mabusa Besala		2	6	0	0	0	0	8
Marimane		10	3	0	0	0	0	13
Mthombo		44	18	0	0	0	0	62
Nqobangolwazi		67	41	2	1	0	0	111

Provinces	Schools						Grand Total
	1	2	3	4	5	6	
Ntsako	66	91	3	4	0	0	164
Reggie Masuku	44	34	5	1	0	1	85
Thuto Pele	51	39	2	1	0	0	93
Tshiawela	36	37	3	1	0	0	77
Grand Total	448	387	23	12	2	1	873

In relation to invalid entries (90), we find that there is a far larger number of timeouts (873) that occurred, i.e. learners choose to provide no input to the service in many more instances rather than provide invalid (wrong) input. We believe that this may be due to the fact that learners were still unfamiliar with the service in the initial phases of the interaction and chose to re-listen to the service again before deciding to make any kind of input.

7.5.2 Coordinator reporting service

The call log analysis showcased in this section covers the calls made to the Coordinator reporting service by school coordinators in 14 schools during the period May 2011 to December 2012. Similar to the Learner feedback service, we keep cognisance of the fact that the pilot period for the schools varies. All analyses were performed on calls made to the service on post-pilot days.

7.5.2.1 Call volume

Across the analysis period, the service had a total of 336 calls. Table 20 illustrates the call distribution per school per month across all provinces. From this we observe that a large number of calls (151) are classified as “Unknown” since the user called from an unregistered mobile phone number, failed to provide their correct PIN, and thus was unable to log into the service. The calls where the school was identified (185 calls across all provinces) are a result of the coordinators using either the phone numbers (mobile/landline) which were pre-registered on the service (thus no login required), or using another phone not registered on the service (login required) but providing the correct PIN assigned to them.

Table 20 illustrates the call distribution across all schools, where we note that across the period of May 2011 to December 2012, Thuto Pele has the highest number of calls with 28% (of the 185 calls) followed by Mabusa Besala with 15% of the calls.

Table 20: Call distribution per school per month (Coordinator reporting service)

Month	School														Grand Total	
	Gauteng		North West		Mpumalanga				Eastern Cape		Free State		Limpopo			Unknown
	Kokotla	Ntsako	Inkonjane	Thuto	Nqoba	Reggie	Mthombo	Mabusa	Cowen	Kwezi	Iketseng	Cedar	Marimane	Tshiwela		
May-11	0	0	-	-	-	-	-	-	-	-	-	-	-	-	10	10
Jun-11	4	11	-	-	-	-	-	-	-	-	-	-	-	-	4	19
Jul-11	0	0	-	-	-	-	-	-	-	-	-	-	-	-	4	4
Aug-11	0	0	-	-	-	-	-	-	-	-	-	-	-	-	8	8
Sep-11	3	0	-	-	-	-	-	-	-	-	-	-	-	-	11	14
Oct-11	0	0	-	-	-	-	-	-	-	-	-	-	-	-	10	10
Nov-11	0	1	12	11	2	-	-	-	-	-	-	-	-	-	9	35
Dec-11	0	0	2	6	0	-	-	-	-	-	-	-	-	-	12	20
Jan-12	0	1	0	2	0	-	-	-	-	-	-	-	-	-	11	14
Feb-12	0	0	10	1	1	3	-	-	-	-	-	-	-	-	7	22
Mar-12	0	0	0	6	0	8	-	-	-	-	-	-	-	-	4	18
Apr-12	0	0	0	8	0	0	-	-	-	-	-	-	-	-	19	27

Month	School														Grand Total	
	Gauteng		North West		Mpumalanga				Eastern Cape		Free State		Limpopo			Unknown
	Kokotla	Ntsako	Inkonjane	Thuto	Nqoba	Reggie	Mthombo	Mabusa	Cowen	Kwezi	Iketseng	Cedar	Marimane	Tshiwela		
May-12	0	0	0	7	0	8	-	-	-	-	-	-	-	-	12	27
Jun-12	0	0	0	1	0	0	-	-	-	-	-	-	-	-	2	3
Jul-12	0	0	0	3	0	0	-	-	-	-	-	-	-	-	12	15
Aug-12	0	0	0	2	0	1	0	3	0	0	-	0	-	-	3	9
Sep-12	0	0	0	0	0	0	2	3	0	1	15	0	2	1	1	25
Oct-12	0	0	0	3	0	1	0	9	1	1	4	0	2	6	10	37
Nov-12	0	0	0	2	0	0	0	13	0	0	0	0	0	2	2	19
Dec-12	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0
Grand Total	7	13	24	52	3	21	2	28	1	2	19	0	4	9	151	336

Figure 19 illustrates the distribution of the 185 calls where the schools were successfully identified. Here we note that the highest number of calls was experienced during the months when pilots and school visits were conducted, e.g. November 2011, February to March 2012, September to November 2012.

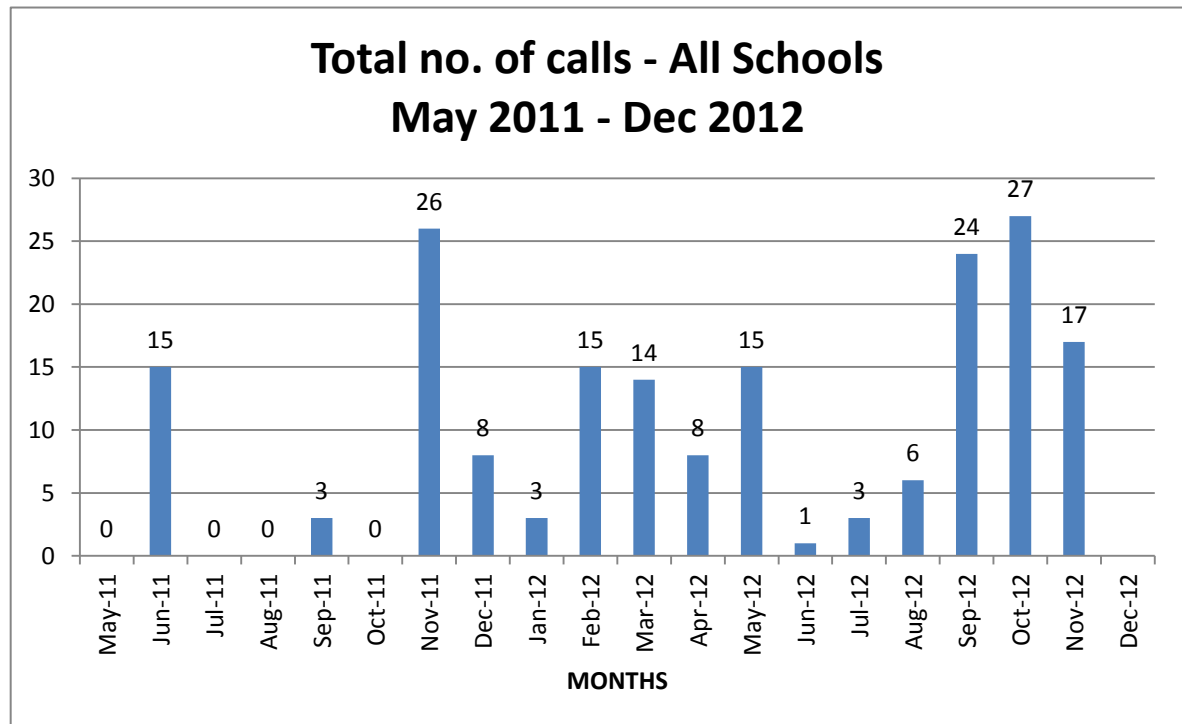


Figure 19: Total no. of calls across all schools

Table 21 shows the average number of calls per week calculated per school. Here we notice that the majority of the schools were calling in less than once a week, with a few schools such as Mabusa Besala (MP), Iketseng (FS), and Thuto Pele (NW) calling in once or more per week.

Table 21: Average calls per week

Month	School													
	Gauteng		North West		Mpumalanga				Eastern Cape		Free State		Limpopo	
	Kokotla	Ntsako	Inkonjane	Thuto	Nqoba	Reggie	Mthombo	Mabusa	Cowen	Kwezi	Iketseng	Cedar	Marimane	Tshiwela
Calls	7	13	24	52	3	21	2	28		2	19	0	4	9
No. of pilot weeks	71.3	71.0	57.3	57.1	56.6	42.3	16.1	15.1	14.3	14.1	13.0	14.6	13.3	13.4
Calls/week	0.1	0.2	0.4	0.9	0.1	0.5	0.1	1.9	0.1	0.1	1.5	0.0	0.3	0.7

7.5.2.2 Call duration

In Figure 20 we observe that some calls have a higher than average (4min 53s) call duration. This is because a portion of the calls (66 calls or 35% of the calls) have very high call durations in the order of 15 min or more (tail end of Figure 20). On analysing these calls further, we found that in some of them, the school coordinators used a single call to report feeding statistics for multiple reporting days by using the “Report for today or another day” option provided by the service thus resulting in higher call duration. We also observe that 48% of the calls lasted between 15 s and 4 min 15 s, while 17% lasted between 4 min 15 s and 14 min 45 s.

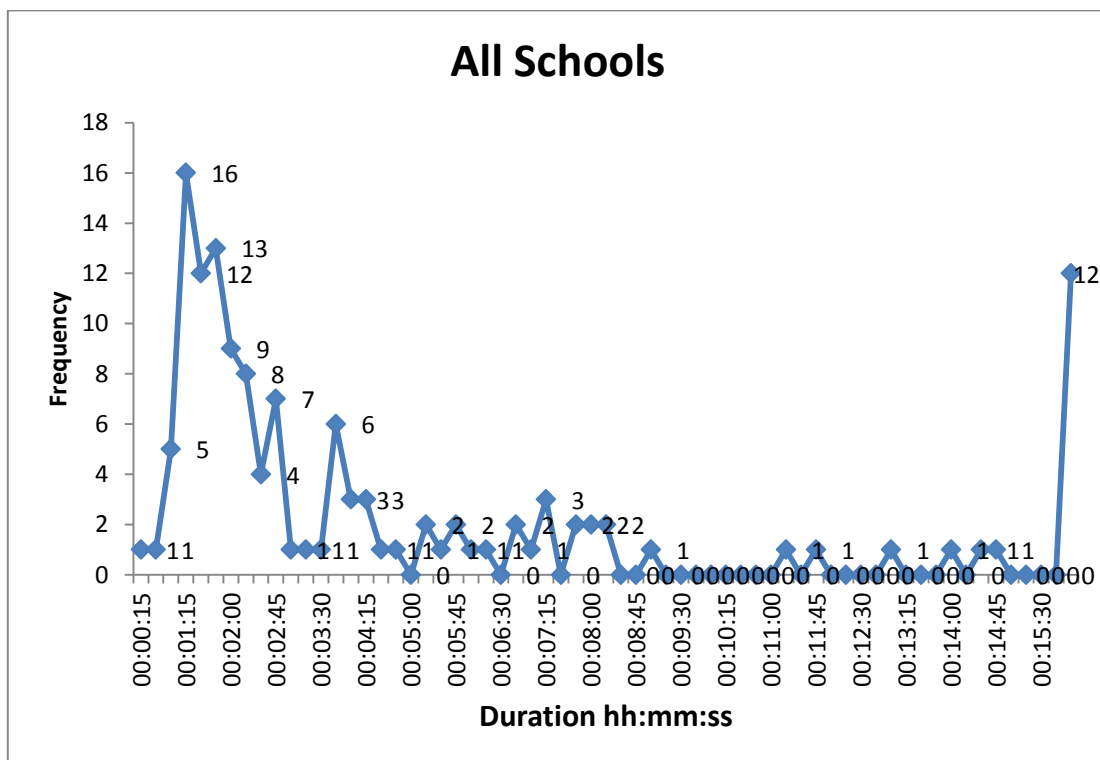


Figure 20: Call duration distribution

7.5.2.3 Final call state

The final call state represents the last dialog state (or step) in the call flow of the service where the user ended the call. Overall this final state gives an indication whether the user was successfully able to complete the task of providing a report about the meals served at their school. It also provides more insight into the possible areas of the call flow where users are dropping off, which might require further attention. Figure 21 illustrates the final call state for calls received across all the schools. We observe that out of 185 calls, 98 calls (52%) ended successfully with coordinators hanging up at “0.25 Leave Exit Message” where coordinators were thanked for their feedback and were given the option to report for another day if they would like to do so.

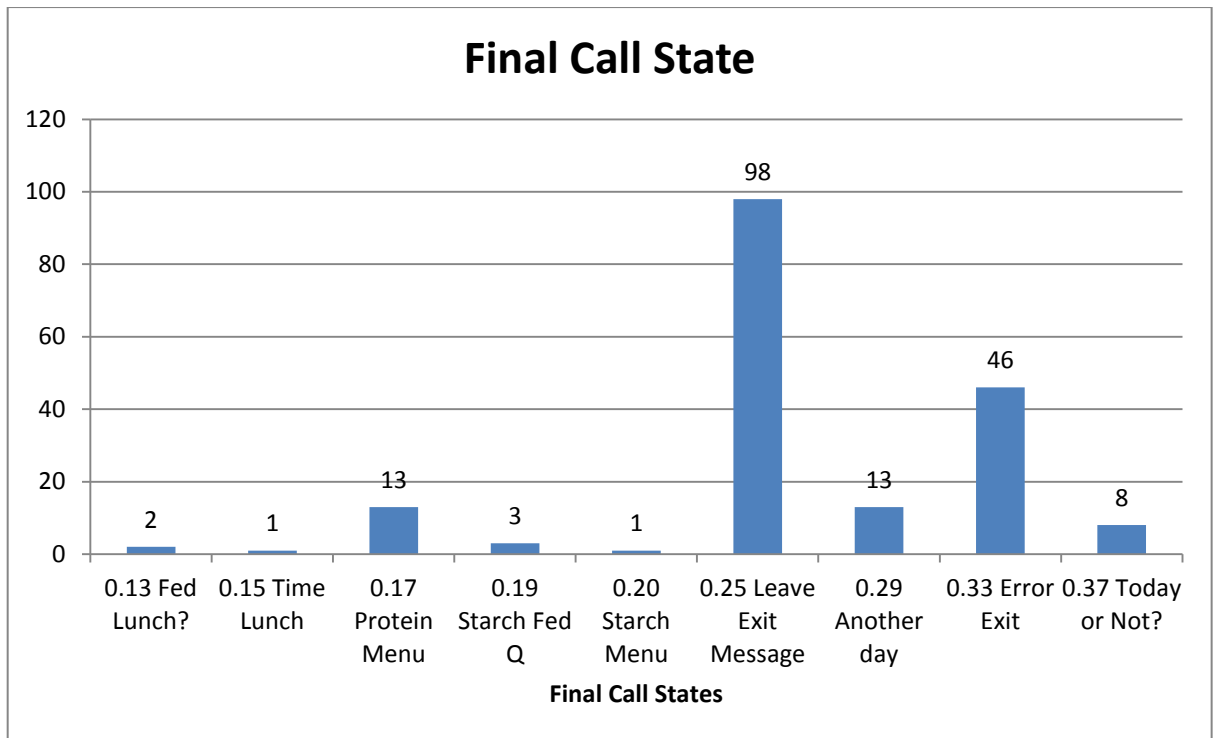


Figure 21: Final call state

Table 22 shows the breakdown of the final call state for each school. We observe that 60% (53% +7%) of calls ended at a dialog state where the school coordinator has successfully gone through all the dialog states for reporting about the school meals, and is either at the state of leaving a message (0.25 Leave Exit Message) or reporting for another day (0.29 Another day). Another 25% (0.33 Error Exit) of the calls are attributed to the case where the user did not make any input or made invalid input, three times in succession, and thus was re-prompted appropriately twice and on the third time the service ended the call after providing them with an option a leave a message for further assistance.

Table 22: Final call state per school

Final Call States	Gauteng		North West		Mpumalanga			Eastern Cape		Free State	Limpop		Freq	%	
	Kokotla	Ntsako	Inkonjane	Thuto Pele	Reggie	Mabusa	Mthombo	Nqoba	Cowen	Kwezi	Iketseng	Marimane			Tshiawela
0.13 Fed Lunch?	0	0	1	1	0	0	0	0	0	0	0	0	0	2	11%
0.15 Time Lunch	0	0	0	1	0	0	0	0	0	0	0	0	0	1	
0.17 Protein Menu	0	0	0	2	0	1	0	0	0	1	1	4	4	13	
0.19 Starch Fed Q	0	0	0	0	0	1	0	0	0	0	2	0	0	3	
0.20 Starch Menu	0	1	0	0	0	0	0	0	0	0	0	0	0	1	
0.25 Leave Exit Message	2	7	17	18	29	22	0	0	0	0	3	0	0	98	53%
0.29 Another day	3	1	0	6	0	1	0	0	0	0	1	0	1	13	7%
0.33 Error Exit	2	4	6	15	1	3	2	2	0	1	6	0	4	46	25%
0.37 Today or Not?	0	0	0	0	0	0	0	1	1	0	6	0	0	8	4%
Grand Total	7	13	24	43	30	28	2	3	1	2	19	4	9	185	

The 151 calls where schools could not be identified since the user called from a phone other than the one pre-registered on the service and was unable to provide the correct PIN, were also analysed as presented in Table 23. We find that the vast majority of the calls (80%) ended at a state where after two tries with an incorrect PIN, the user is asked to leave their name and number for assistance (0.12 Leave Name & Number).

Table 23: Final call state – "Unknown" schools

Final Call States	Frequency	%
0.1 CallerId Intro	4	3%
0.10 Login	25	17%
0.12 Leave Name & Number	121	80%
0.37 Today or Not?	1	1%
Grand Total	151	

7.5.2.4 Invalid entries

From the 185 calls with identified schools, we analysed the invalid entries and found that there were 41 calls in total with invalid entries (22% of total calls). Of these we find that 32 calls (17%) had only one invalid entry, which implies that after the first invalid entry users were able to successfully provide correct input after being prompted by the service, and thus proceed with the call. Table 24 illustrates the occurrence of invalid entries per school.

Table 24: Invalid entries per school

Provinces	Schools	No. of invalid entries/call			Total
		1	2	3	
Gauteng	Kokotla	2	0	0	2
	Ntsako	1	1	0	2
North West	Inkonjane	3	1	0	4
	Thuto Pele	12	3	2	17
Mpumalanga	Mabusa Besala	2	1	0	3
	Mthombo	2	0	0	2
	Nqobangolwazi	1	0	0	1

Provinces	Schools	No. of invalid entries/call			Total
		1	2	3	
	Reggie Masuku	3	1	0	4
Eastern Cape	Cowen	0	0	0	
	Kwezilomso	2	0	0	2
Free State	Cedar	0	0	0	
	Iketseng	1	0	0	1
Limpopo	Marimane	0	0	0	
	Tshiwela	3	0	0	3
Grand Total		32	7	2	41

7.5.2.5 Timeouts

Table 25 shows the breakdown of the calls with timeouts found over the analysis period. Of the total 185 calls, there were 71 calls with timeouts, where the majority of calls with timeouts (69) were with one and two timeouts (37% calls). We note that the number of timeouts is higher than the number of invalid entries; here it seems again that users would rather provide no input and listen to the prompt again, than provide wrong input.

Table 25: Timeouts per school

Provinces	Schools	No. of timeouts/call			Total
		1	2	3	
Gauteng	Kokotla	2	1	0	3
	Ntsako	8	0	0	8
North West	Inkonjane	9	0	0	9
	Thuto Pele	9	6	0	15
Mpumalanga	Mabusa Besala	6	3	0	9
	Mthombo	0	0	0	
	Nqobangolwazi	0	0	0	
	Reggie Masuku	6	0	0	6

Provinces	Schools	No. of timeouts/call			Total
		1	2	3	
Eastern Cape	Cowen	0	0	0	
	Kwezilomso	0	1	0	1
Free State	Cedar	0	0	0	
	Iketseng	11	1	2	14
Limpopo	Marimane	0	0	0	
	Tshiwela	3	3	0	6
Grand Total		54	15	2	71

7.5.3 Discussion

From the analysis of the two services in the preceding sections we observe that the Learner feedback service had a much higher usage and uptake than the Coordinator reporting service.

However, we do note that the Learner feedback service's usage also saw a usage decline in terms in schools where the service had been running for longer, e.g. Gauteng, North West, and the first set of Mpumalanga schools. The highest usage was in the months July 2011, November 2011, March 2012, August 2012, and September 2012, when the service was piloted in schools or when site visits were conducted by the team. Overall, Mpumalanga (31%) and Gauteng (29%) have the highest usage, followed by North West (21%), keeping in mind that these schools were only piloted in phase 1 and phase 2 where the service had a longer running time (only two schools in Mpumalanga were piloted in phase 3). Thus we further analysed the average calls per week from a school and found that most schools had between 5-9 calls/week, with a few in the 2-4 calls/week range, and a high of 16.8 calls/week which is attributed to a short pilot period (September to December 2012).

In terms of overall language usage, we observe that although English (44%) may have been individually the most preferred language, collectively African languages were chosen for 54% of the calls, making a strong case for local language service delivery. In particular, the provincial language analysis reveals that only for Gauteng (57%) and North West (46%), English was the most preferred language with Sepedi (27%) and Setswana (40%) being the second and third most preferred languages respectively. In the other provinces isiZulu was highest in Mpumalanga at 46% (Eng 39%), isiXhosa was most preferred in Eastern Cape at 54% (Eng 37%), Free State province had Sesotho as most preferred at 55% (Eng 40%), and Limpopo had Tshivenda with 41% (Eng only 28%).

The final call state analysis also revealed that learners were successfully able to use the service to report on their school meals (55%), but a significant 29% of the calls were hung up in first few stages of the call (introduction or language menu) which may indicate prank calling or learners being unsure in using the service. There is a relatively low number (3%) of invalid entries (wrong input) by the learners, which indicates that learners are comfortable using speech and telephone-based technologies. In comparison, the number of timeouts (no input) is much higher (14%) but we find that only 1% of these were unrecoverable (ended in error exit). 13% of users recovered had a timeout within the call but were able to continue the call through the error recovery prompts provided.

For the Coordinator reporting service we note that there was no, or very little, usage in the schools that were piloted in phases 1 and 2 (with the exception of Thuto Pele and occasionally Reggie Masuku). The phase 3 schools saw usage being gradually tapered down in the months post-piloting, with only two schools (Mabusa Besala and Ikeketseng) being more active.

The final call state analysis indicates that at least 60% of the calls ended successfully with 25% of calls resulting in error exits, which is notably higher than the Learner feedback service, as well as invalid entries (22% of total calls though 21% recovered within the call) and timeouts (38% of which 37% recovered). From a quantitative perspective one of the other major challenges experienced was the high number of calls with users using unregistered phone numbers who were unable to log into the service as they were not able to provide the correct PIN (see discussion under 7.6.2).

7.6 User experience evaluation

7.6.1 Learner feedback service evaluation

7.6.1.1 Background

Table 26 summarises the statistics of the evaluation of the Learner feedback service. The evaluations took place over a nine-month period starting in February 2012 and ending in November 2012. In all instances the service had been implemented at the pilot site for at least six weeks, before being evaluated.

Table 26: Evaluations of the learner feedback service

School	No. of participants	Province	No. of focus groups	Evaluation date
Kokotla	40	Gauteng	2 user and 2 non-user focus groups	11 February 2012 17 March 2012
Ntsako	33	Gauteng	2 user and 2 non-user focus	18 February 2012

School	No. of participants	Province	No. of focus groups	Evaluation date
			groups	10 March 2012 24 March 2012
Inkonjane	21	North West	1 user and 1 non-user focus group	27 May 2012
Thuto Pele	21	North West	1 user and 1 non-user focus group	26 May 2012
Nqobangolwazi	18	Mpumalanga	1 user and 1 non-user focus group	2 May 2012
Reggie Masuku	18	Mpumalanga	1 user and 1 non-user focus group	13 May 2012
Mthombo¹³	16	Mpumalanga	1 focus group consisting of users and non-users	18 October 2012
Cowen	21	Eastern Cape	1 focus group consisting of users and non-users	31 October 2012
Marimane	21	Limpopo	1 focus group consisting of users and non-users	7 November 2012

7.6.1.2 Objectives

The purpose of these evaluations was to determine the following related to the Lwazi II Learner feedback service as a feedback mechanism for the NSNP:

- The learners' experiences of the Lwazi Learner feedback service.
- The uptake of the Lwazi II system by Grade 8 and 9 learners.
- Reasons for no or limited use of the system by Grade 8 to 9 learners.

¹³ Only 3 phase 3 schools were selected (1 each from 3 provinces) for evaluation due to time constraints

- Learners’ perspectives on the potential impact of using the Lwazi system on the efficiency of the NSNP.
- Whether or not learners had had a prior demonstration of the system.
- The efficiency of the piloting methodology in ensuring frequent and consistent use of the system by the learners.

7.6.1.3 Research methodology

7.6.1.3.1 Data collection method used

The user experience study used only one method of data collection, namely focus group discussions. The qualitative data thus obtained, complements the quantitative data presented in the call log analysis section above.

According to Leedy and Ormrod (2005), the most commonly used type of qualitative methodology for gathering valuable user data and an acceptable tool for a multitude of groups, is the focus group discussion (FGD). FGDs entail the use of a researcher as a facilitator who prompts discussion among the group members. The method is used to make sure that, unlike in structured interviews where closed questions might reflect the interviewer's bias and opinions rather than those of the interviewee, open-ended questions are used as the participants’ views are most valuable.

FGDs allow for a carefully planned discussion among the members of a group that is facilitated or moderated by a skilled interviewer. Participants are selected because they have certain characteristics in common that relate to the topic of the focus group. Although the FGD may take an hour or more to complete, it provides opportunities to collect data from a group of people while simultaneously observing the interaction among the participants.

A total of 19 FGDs were held in nine schools across all three phases.

7.6.1.3.2 Selection of respondents and constitution of groups

In phases 1 and 2 the selection of the learners to participate in the FGDs was done two to three weeks before the discussion date. For phase 3, respondents were selected or asked to volunteer the day before the discussion. Two members from the HLT monitoring and evaluation team visited the selected schools on two consecutive days. On the first day they handed out consent forms (see annexures B and C) to obtain consent from learners and their guardians for participation in the FGDs. On the second day the team met with the learners and selected all learners with fully completed consent forms for the discussion. Learners with unsigned forms were not included in the evaluations.

In phase 1, four FGDs were held per school – two with learners who had used the service (so-called “user groups”) and two with users who had not used the service (“non-user groups”). In phase 2, this was brought down to two groups per school (one user group and one non-user group). In phase 3, only one FGD was held per school, consisting of users and

non-users, as well as ambassadors and non-ambassadors. In phases 1 and 2, all FGDs were held on Saturdays to ensure minimal interruptions to academic activities during school hours. In phase 3, the evaluations took place in the afternoon, after school hours.

7.6.1.3.3 FGD procedure

To enable the learners to relax, the FGDs started off with the playing of a game of UNO. In general, the learners were energetic and interacted well with each other. Shy or quiet learners were encouraged to talk by going around the groups and asking learners to respond to questions one at a time. This ensured responses and ideas from all learners.

An interview schedule (see annexure D) was used to direct the FGDs. The interview schedule guided the respondents to speak about their favourite foods, then about food at their school and finally about the Lwazi service.

7.6.1.4 Findings and recommendations

The following represents a summary of findings and recommendations from the focus group discussions, across all three phases. The data was integrated and categorised and has been supplemented with quotes below to elucidate the analysis.

7.6.1.4.1 NSNP

Learners' experience of the NSNP

Generally learners were satisfied with the meals received at school. There were specific comments on the quality, food preparation etc. of the meals which will be covered below.

Quantity of meals provided

Some learners commented that the quantity of food was not always sufficient – especially when their favourite meals were prepared. On other days, however, learners would not eat the food provided and it would be thrown away.

“We eat ... the meal I don't like is Tuesday. We eat macaroni with Soya mince. We don't eat that. It rots when they cook it because we don't eat.”

Learners also claimed that the VFHs and teachers took food home and would only provide extra helpings to certain learners.

Food preparation

Some learners indicated that the food was so badly prepared that they intended to strike. Some commented that the meals were not always cooked well or were even old or busy rotting.

“The main fact is that the school is going to riot!! We don't care.”

Some learners were of the opinion that the food was no longer tasty due to the change in the time of serving meals. The volunteer food handlers (VFHs) no longer have enough time to prepare food because they arrive at 08:00 and the food has to be ready by 10:00. With receiving their meals so early, learners commented that they were hungry and still needed to concentrate on school work for the rest of the day.

“The worst thing is that they changed break time to be earlier. We get hungry because we spend a lot of time at school. ... When we eat lunch, we are still full from breakfast. ... At 10 it’s breakfast, and then soon it’s lunch. We go home hungry. We do not concentrate during the afternoon classes, we sleep.”

Learners mentioned that when they got the announcement from an official at DBE that the food time table had changed, they were surprised that they had not been consulted.

Learners also commented on the state of some of the kitchens, indicating that they were very dirty.

Complaint channel

Some learners felt that teachers/coordinators were not approachable and that they could not raise their complaints related to meals with the teachers/coordinators. When asked how and to whom they usually addressed complaints, learners mentioned that there was a set structure where learners inform the class representatives, who then inform the teacher, who then informs the deputy principal.

“They report to us [class reps] and we tell class teacher, the teacher tells Mr XYZ.”

Other issues

Some of the boys commented that the girls in their schools receive **preferential treatment** and are given better food than them.

“The same with fish. The boys eat a different brand from girls. Also the bean soup, the boys’ soup is brown but the girls’ soup is lighter.”

Some learners were a bit dissatisfied at having to wash their own cups and dishes after being served their meals.

The VFHs are parents who volunteer to cook meals at the schools. Learners felt that they did not want to call the system in the presence of learners who are related to VFHs as they may be **“punished”** for complaining about the meals.

Learners mentioned that there is a **stigma** attached to receiving meals at school, these also deterred some learners from using the system.

“I think many learners are not interested in the feeding scheme. I was thinking that you motivate them to eat. I think that is what will motivate them to call. They reason they don’t call is that they don’t eat so they are not interested. They say the food is for poor people.”

7.6.1.4.2 The Lwazi system

Learners' exposure to the piloting

Most learners had either participated in or were aware of the system that was piloted at their schools.

"[With the] Lwazi system we were supposed to complain about good or bad food. So at first the food was tasty but now it has changed to be bad."

In some of the phase 3 schools, the ambassadors had informed their classmates about the system and had encouraged other learners to use the system. However, it was clear that some learners did not realise that "Lwazi", the "system" and the "number to call" were all part of the same thing.

"I told them that they should call the number after eating. I showed them the number to call. We put a poster in class. I demonstrated with a phone."

"In class they told us about a number that we call to say how the food was. They did not tell us about Lwazi."

"In our class they put the poster in front and said we should buzz the number. And then we will talk to a certain lady. We will tell her if the food was bad or good."

Learners' experience of the system

Generally, the learners mentioned that they found the system interesting and easy to use, and liked it.

"... because the system is very easy to use."

Most learners remembered how the system worked. A few learners even mentioned that they had been using the system continuously since being introduced to it and others said that they had used it a few times after the pilot.

Some were afraid to call.

"What I heard is that when you call your number shows and then the information gets back to the school."

Some indicated that the system had not called them back after they had given it a missed call.

Language technology is a new concept to learners. Although interesting, it was an experience that was at times not fully understood.

Learners' access to the system

The majority of learners indicated that they have their own cell phones.

Some learners only used the system when it was piloted, thereafter forgot about it. Other learners mentioned that they had not used the system as they were not allowed to bring cell phones to school (to call when they saw the posters) and had forgotten the number by the time they got home. (The Lwazi team did mention that the lines were open 24 hours a day, 7 days a week; therefore, learners (if they don't forget) can call the system any time.)

Misunderstandings

When to use the system

A number of learners knew about the system but did not call because they had nothing to complain about. There seemed to be a general perception that the Lwazi system was a complaints line.

Some learners thought that they were supposed to call the system only when the food was not good. They did not realise that they could call the system when the food was good or bad.

Some thought that the system could only be used at school.

Cost

There also seemed to be misunderstandings relating to the costs involved.

"In my class I heard them saying that the number uses your airtime."

How the system works?

Some learners mentioned that they were confused about the use of the system; and were not calling as often as they could.

Some learners were under the impression that a "please call me" could be used for the system to call them back. Even though there is currently no "please call me" functionality on the system, some learners even claimed that the system had called them back on a "please call me".

The role of CSIR/DBE

Learners were at times confused about the role of the CSIR versus DBE and who would solve the issues they raised. Some learners thought that the CSIR team was the same as the DBE team. There was also a misperception about the relationship between the Lwazi system and the DBE. Some learners indicated that nothing had changed since the Lwazi system had been implemented. Learners wanted to see immediate change (i.e. call today, and get results tomorrow).

After the CSIR explained that change cannot take place so quickly¹⁴, learners felt a bit more comfortable to use the system and to share with their fellow learners.

Learners questioned what would happen if the DBE found messages where learners had complained about the food they received at school.

Who is Mama Nandi?

Some learners asked where “Mama Nandi” is. The facilitators explained that “Mama Nandi” was a fictional person (persona) who had been created to engender trust in the system. “Mama Nandi” was at times misunderstood. At one school, an announcement for volunteers to participate in the focus group discussions was made at the school assembly. The principal apparently asked: “Who called Mama Nandi?” – learners thought that Mama Nandi had “told on them”. In this case, this led to the Lwazi system being mistrusted.

Tone and language of the system

Most learners interacted with the system in their home languages, while a few indicated that they had used English. Those who chose English said they did it “for fun” while others said they did it so that Mama Nandi could understand them. Some learners indicated that they preferred to select English to interact with the system because this was the language they used to communicate with people who did not understand their language, or with people whom they did not know.

“I won’t use English because it is not my mother tongue.”

“I will use English because if you do not understand Tsonga, you will at least understand English.”

The messages were, however, left in their home languages. All prompts in all languages assessed were clear. Learners did not suggest any changes to the languages or the voices.

Mistrust of/uncertainty about the system

There were mixed feelings about trusting the system. Some learners said that they trusted the system because it did not ask for a name and that the voice was friendly enough for them to open up to.

“I think her voice is ok because she is open, you are able to speak to her about all the things you want to tell her. When you hear her voice you say all the things you want to say.”

Others were uncertain about the anonymity of the system.

¹⁴ The metaphor of a Ferrari (the Lwazi system) without fuel (resources to enable the system to operate effectively) was used to explain this to learners.

Usage model

At times learners would call the system while sitting in a group. When doing this, learners would use one phone and make one call, and provide feedback to another learner holding a phone in the background.

“Like the way we are sitting now [in a circle], you will find that this side is quiet and answers nothing, while there ones will report [meaning the others in the group].”

Non-use of the system

Some learners who participated in the focus group discussions had never used the Lwazi system. They had either not been told about the system by classmates or ambassadors or they had not participated in the actual piloting. Unfortunately, we found that there were also ambassadors in our phase 3 schools who had not used the system. This was problematic as they had been selected to introduce the system to their fellow learners.

There also seemed to be inaccurate information provided to learners that the system used lots of airtime. It was explained that in order to give the system a missed call, airtime was required; however, this would not cost learners anything. Many learners did not call the system as they did not have airtime to give the system a missed call.

Some learners had never heard of the system, while others did not think the project was serious.

As indicated under accessibility above, learners were not allowed to take phones to school, therefore, forgot about using the system, since the posters were at school.

7.6.1.4.3 Challenges

Learners felt that the system was too expensive to call. Some ambassadors claimed that they had called the system and that it had charged them; so it was too expensive.

Some learners were afraid to call the system due to claims that teachers would accuse them of whistleblowing.

7.6.1.4.4 Recommendations

Visibility of systems

Learners requested that there be an increase in the visibility of systems. This could be in the form of additional marketing materials or more regular visits by the piloting team.

Access to phones

Learners felt that there should be an increase in phone lines – that public phones should be installed.

Additional piloting

Since many learners had not heard of the system, they recommended that the system should be piloted again and explained to learners in more detail. Learners should also be provided with increased opportunities to test the system during piloting.

The learners felt that it more site visits to schools by the piloting team, might help to increase uptake.

Effectiveness of the system

Learners requested the CSIR/DBE to ensure that the system is fully functional at all times. If there are problems with the system (i.e. does not call back), this deters learners from using the system again.

Additional marketing

Learners requested that additional brochures be distributed which they may take home. This would help to remind them to use the system when they got home.

“Please call me”

Learners suggested that “please call me” be implemented as a way to access the system (instead of a “missed call”) as this would not require any airtime; which was an issue for them.

Feedback on calls

As indicated above, learners would like to know whether they would receive feedback or results, is they used the system.

7.6.2 Coordinator reporting service evaluation

7.6.2.1 Objectives

The objectives of the evaluation were –

- to evaluate the usability of the Coordinator service;
- to learn about reasons why they would or would not use the service;
- to investigate whether if the service adds value to the day-to-day management of the NSNP at school level; and
- to investigate the scalability of the service.

7.6.2.2 Research methodology

In phase 1, set questions were drafted for the interview. These were divided into sub-categories entitled: questionnaire & interview details; respondent’s information; demographics; introduction to the NSNP; Lwazi II pilot project and use of the system. In phases 2 and 3, an interview schedule with four open-ended questions (attached as

annexure E) was used to conduct key informant interviews with the sampled school coordinators. Schools, and therefore coordinators, were conveniently selected.

In phase 3, the interviews took place on the second day of the evaluation at schools. One of the HLT team members met with the school coordinator and conducted an individual interview, probing for additional information, where required. Not all school coordinators were interviewed.

7.6.2.3 Findings and recommendations

The following represents a consolidated list of findings and recommendations from across all three phases of evaluation:

- The CSIR should present a training workshop to DBE staff to ensure wider scope of understanding of the benefits and use of the system.
- The CSIR should send SMS reminders to the coordinators to use the system at 13:20 and not at 15:00.
- The coordinators were more comfortable using the system once per week, instead of daily.
- The coordinators should be allowed to provide a ball-park figure of the learners fed (because counting takes too much time).
- There should be more regular interaction between school coordinators and the DBE on a way forward for using the system more effectively.
- There should be shared responsibility for using the system (additional school staff should be made responsible for reporting information – not only the coordinators).
- Staff at all levels of government should be able to access the information on the Lwazi service at all times.
- The service should include questions regarding utensils and facilities which seem to be a big challenge at schools.
- Some coordinators found that the service was easy to use and did not take too much of their time, however, others found that it took too much time reporting daily.
- The coordinators were generally satisfied with the language, structure and questions asked by the service and found that the pace and response time was adequate.
- The general expectation amongst the coordinators was that the inputs and messages on the system would be heard and addressed by the Department of Basic Education.

7.6.3 Web monitoring interface evaluation

This interface was only evaluated in phase 3, as district officials were selected to evaluate the interface.

7.6.3.1 Objectives

The purpose of this evaluation was –

- to evaluate the usability of the web monitoring service;
- to investigate whether the service adds value to the day-to-day management of the NSNP; and
- to investigate the scalability of the service.

7.6.3.2 Research methodology

To evaluate the web monitoring interface, NSNP district officials were sent a questionnaire with four questions for them to answer online in the comfort of their offices. The same questionnaire was used as with the Coordinator reporting service evaluation. Unfortunately only one of three district officials responded.

7.6.3.3 Findings and recommendations

The respondent indicated that –

- the service was informative and helpful in terms of identifying problems at schools;
- the service could provide an opportunity to resolve issues sooner;
- if used daily, the service could enable officials to write monthly reports for the district and narrative reports for the provincial department; and
- the service should be rolled out because it was useful for compiling a report and assisted the Department to obtain direct information from learners and coordinators.

It may thus be concluded that the web monitoring interface provides an instant and easy way for district and provincial DBE officials to access information on schools that they otherwise would only be able to obtain through visits to the schools. The added usefulness of this interface, as highlighted by the district manager who responded, is that it enables users to compile monthly reports.

7.6.4 Conclusions

There was a mixture of responses across all users and non-users of the DBE NSNP Lwazi services. On the Learner feedback application, we learnt that there were many misunderstandings on what the purpose of the system was, including its benefits. The general impression was that if problems were reported, these would be solved immediately.

Despite positive feedback there was little enthusiasm among school coordinators for the use of the system. Work load was indicated as the main reason for not using the service optimally. As with the Learner feedback service, the coordinators also felt that they would prefer to see tangible results on what they had reported.

At the district level, department duties seemed to take priority over external projects such as Lwazi. Even though on pilot days we had a full and well-represented delegation from the district office, the use of the system was not optimal, and only one district official (out of three respondents selected) completed the evaluation questionnaire.

This indicates that future roll-outs of the system would need to be supported through an extensive marketing campaign, visible support from the national DBE, and the implementation of a process to deal with issues raised through the system, in an effective and speedy manner.

8. Afrivet services

8.1 Background

Afrivet Training Services (<http://www.afrivet.co.za>) is a South African primary animal health care (PAHC) and training services provider whose focus is on servicing rural areas in South Africa. Afrivet works with various role players such as state veterinarians (vets), rural private vets (as opposed to urban vets who work with domesticated animals), livestock farmers, community animal health workers, etc., as highlighted in Figure 22 below (approximate user population sizes given in brackets).

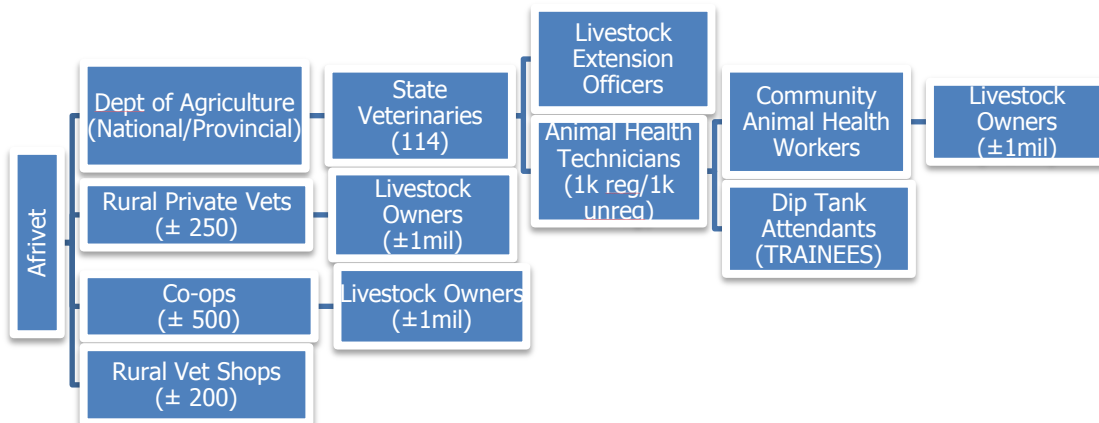


Figure 22: Overview of PAHC role players

It is estimated that a loss of approximately R3 billion occurred in 2010/11, as a result of animal diseases (Afrivet Training Services, personal communication, 5 March 2012). Prevention depends on accurate and timely information on disease identification reaching the various role players especially the rural veterinarians. The need for information is two-way; 1) disease surveillance data is required for early detection of diseases to prevent loss of livestock, and 2) information on disease outbreaks, notifications for new products and training need to be disseminated in the field to the various role players (Afrivet Training Services, personal communication, 5 March 2012).

Our work with Afrivet Training Services commenced in 2012 with a number of in-depth joint requirements planning sessions to obtain a thorough understanding of the PAHC environment, the needs and challenges, potential applications and the specifications for such applications.

The joint requirements planning sessions led to the development of a detailed system architecture design which guided the design and development of the various applications, and the accompanying support interfaces thereof, as described in section 8.2.1 below.

8.2 Design

8.2.1 Afrivet System Overview

The Afrivet service uses a combination of inbound voice-based telephone services and an outbound alerting service (SMS, email, voice) to provide a holistic solution for Afrivet to communicate with the various PAHC role-players. Figure 23 provides an overview of the Afrivet system architecture, where the top and bottom layers represent the user interface layers of the voice services, and web-based services (registration, outbound alerting, transcriber dashboards, etc.). The logic and database layers in turn depict the various inter-relationships amongst the services and the corresponding high-level back-end databases required.

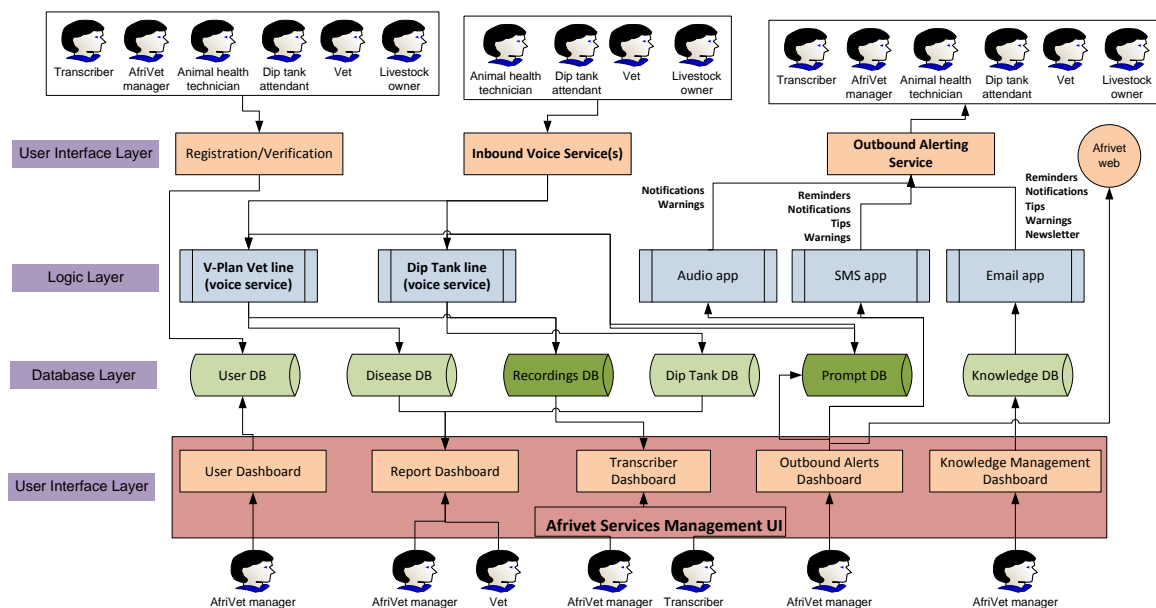


Figure 23: Afrivet system architecture

In terms of the voice-based services, we designed two inbound services: 1) the V-Plan Vet line for rural vets, and 2) the Dip tank line for dip tank attendants (DTAs), as described below.

8.2.2 V-Plan Vet line

The V-Plan Vet line, available in English and Afrikaans, allows private rural vets to call in and report on disease incidences encountered in the field, through answering a series of questions around the incidence (Figure 24). The interaction is purely speech-based, as the context of use typically involves vets calling the service right after a client callout whilst driving, using their mobile phone in noisy environments (farm area with animals, and rural roads). Communication amongst the rural veterinarians and with Afrivet for issues such as disease outbreaks typically occurs through phone calls, with occasional emails. Currently, Afrivet uses a paper-based disease reporting system where vets report on a monthly basis on diseases they have encountered in the field. Afrivet finds that only 15-20% of the vets submit these reports; most are faxed and a handful are emailed; administrative burden being identified as the primary cause for the low response rate. Although this provides some information to Afrivet on disease identification and treatment, it does not provide an early warning system for flagging disease outbreaks.

Sys 0.1: Welcome to the V-Plan vet line. Let's start with the first report on a disease or nutrition problem. How many farms is this related to? Say the number, and then press any key. <tone>
User: 3 farms
Sys 0.2: What type of animal was affected? <tone>
User: Cattle
Sys 0.3: And how many of them were affected? <tone>
User: 12
Sys 0.4: What was your diagnosis or main observations? <tone>
User: Foot and mouth disease
Sys 0.5: Please provide more background information. Say something about the history, outbreaks, climatic conditions, etcetera. Press any key when done.<tone>
User: Second time in a month that herds in this area have been affected.
Sys 0.6: Thanks! If you want to tell us more about the meal, leave your message after the tone and press hash when you're done. Or you can now hang up. <Tone>
User: The food was not well-cooked..
Sys 0.7: Ok. To leave another report, stay on the line. Otherwise you can end the call now.
User: (hang up)

Figure 24: A sample V-plan Vet line interaction

Our initial version of the system was designed to have specific questions that were required by Afrivet, e.g. number of animals affected/dead, but vets found this to be cumbersome, with some even trying to capture multiple disease incidences in one report. As a small experiment, we also created two versions of the system where, at the end of an audio recording, the system either moves on by “pressing the hash key” (coached in the first two prompts only), or through silence detection. Half the users were directed to the silence detection version and the others to the “press hash” version. For the latter, we found that some users forgot to press hash in the prompts when they were not explicitly told to do so (third prompt onwards), and said the system was taking too long to move on (it would move on after a “no-key press” timeout). For the silence detection version as well, vets wanted a shorter silence detection threshold.

There were also feature suggestions such as receiving a case reminder, so vets may follow up on the disease cases they have reported. Overall it was found that the vets were able to easily use the system, and said it would especially be useful if the system could replace the paper-based reporting and also include livestock owners (their clients) in the reporting. Notwithstanding, we found that vets were very time sensitive. Thus the subsequent revised call flow was streamlined significantly, and was launched again in October 2012.

8.2.3 Dip tank line

The second inbound voice service is the Dip tank line in isiZulu and English, which allows dip tank attendants to call in to report on dipping activity and problems at their dip tanks (product used, product required, disease observed, etc.). It uses a combination of DTMF and speech input prompts. A dip tank attendant (DTA) is a communal livestock owner, selected by the community as the caretaker of the local dip tank where animals are taken for regular dipping sessions (varies seasonally from weekly to monthly) and general check-ups/vaccinations. Afrivet works on a regular basis with some DTAs in a few communities, where monthly meetings are held to gather information on dip tank operations (products required, problems encountered, diseases observed).

The dip tank line was introduced to 21 DTAs in a focus group in Siyaphambili (KZN), with follow-up individual interviews with 8 DTAs, where all the DTAs preferred to use the line in isiZulu. The first version of the line was more DTMF-based with some natural speech input for leaving messages. DTAs felt that they liked the speech input better (reasons cited were that it was easier, and also allowed them to leave more information). Overall DTAs interacted well with the system and were keen to use the line because it would be free for them to leave messages for Afrivet (a call-back mechanism) and also because it was provided in their language.

Based on this feedback, a consultative process was again held with Afrivet, where we re-designed the dip tank line to allow more of the questions to be answered via speech input and used DTMF minimally where required, e.g. call branching. We also provided more examples in the prompts for the first five calls per user, to acquaint them with the line. Subsequently, the revised dip tank line was piloted on 18 September 2012 with the DTAs in Siyaphambili.

8.2.4 Web management interfaces

The web management interfaces provide Afrivet with a web-based graphical user interface (GUI) channel to monitor and manage the incoming information from the vet and dip tank lines. Transcribers extract data from the incoming calls through a customized web transcription interface (Figure 25). The speech data collected will serve to bootstrap speech recognition development in the next project phase, as it provides us with real in-field contextual data collected in the environments that users would call from. The transcribed information is channelled to a report management web interface, where Afrivet monitors

the incoming disease reports. Registration of users and user management is managed internally (by CSIR and Afrivet) through the web interface. The incoming calls and transcribed data are channelled to multiple databases which link to the appropriate “dashboards” (e.g. service activity levels, registration requests, and assistance required requests), with which Afrivet and transcribers interact via the web.

Based on the usage of the web interfaces during the first pilots of the Vet line and the Dip tank line, the web interfaces were incrementally but extensively revised to ensure optimal user experience for Afrivet and the transcribers in using the service. From the perspective of the back-end operations, we found that it was difficult for transcribers without experience in the PAHC field, e.g. disease and medication name, to extract reports from the incoming calls. Thus, we subsequently used transcribers to just transcribe the messages, followed by Afrivet extracting the actual report content and quality checking the transcriptions.

Transcriber Dashboard

[Home](#) | [Logout](#)



		Call ID	Caller	App	Date	Lang	Duration	Owner	Status	Last change	Comment
<input type="radio"/>		1574	Matthew Carter	Dip	20 Nov 08:31	English	0:00:10	wpienaar	Unsure	20 Nov 10:10 (wpienaar)	
<input type="radio"/>		1572	Matthew Carter	Dip	20 Nov 08:28	English	0:00:03	wpienaar	Unsure	20 Nov 11:12 (wpienaar)	
<input type="radio"/>		1569	Sibaya Shabalala	Dip	19 Nov 17:58	IsiZulu	0:00:13	smabaso	Unsure	26 Nov 08:14 (smabaso)	
<input type="radio"/>		1547	Nontobeko Judith Mthethwa	Dip	11 Nov 08:32	IsiZulu	0:00:43	smabaso	Unsure	12 Nov 11:46 (smabaso)	

Figure 25: Web transcription interface

8.2.5 Outbound alerting service

The outbound alerting service currently allows Afrivet to send out emails and SMSes instantly to vets and dip tank attendants for any real-time notifications (Figure 26). The SMSes and/or emails can be sent out in multiple languages (English, Afrikaans, isiZulu) and are directed to the recipients based on their registered language and communication preferences. The goal is to afford Afrivet a one-stop place from which to disseminate the information being received from different users, to the multiple PAHC role-players. The longer term vision for this service is to have the capability to send out various types of custom, pre-scheduled, and automated outgoing messages, e.g. call reminders, manager and transcriber notifications for new calls, outbreak warnings, tips, and newsletters, based on user communication preferences gathered at registration. With this in mind, in the design, we focussed on capabilities to customize (e.g. via SMS, email and/or voice; send only to cattle owners), localize (disease outbreaks information sent only to affected provinces), and pre-schedule (set up reminders, seasonal tips beforehand) such messages.

Outgoing message

[Home](#) | [Logout](#)

Create an outgoing message below

To:

User group:

- Vets
- Dip tank attendants

Province:

- All
- Eastern Cape
- Free State
- Gauteng
- KwaZulu-Natal
- Limpopo
- Mpumalanga
- North West
- Northern Cape
- Western Cape

Animal type:

- All
- Cattle
- Sheep
- Goats
- Other

Recipients:

- Wikus Pienaar (V-NW)
- Ane Bekker (V-NW)
- Pieter Jansen van Vuuren (V-EC)
- Nico Degenaar (V-WC)
- Francois Andre van Niekerk (V-EC)
- Dr Dries Lessing (V-FS)
- Peter Oberem (V-GP)
- D'Wall Hauptfleisch (V-FS)
- Danie van Zyl (V-FS)
- Hardus Pieters (V-MP)
- Pieter Nel (V-NW)
- Murray Smith (V-FS)
- Jurie Kritzinger (V-NW)
- Dr Nienke / Dr Johann van Hasselt / Blaaw (V-FS)
- Louis Van Jaarsveld (V-MP)
- Stephen Wessels (V-FS)
- Hanre Bredenkamp (V-GP)
- Chris Nel (V-FS)

Message:

Name:

Name to help identify message (eg. "Weekly Vet line reminder")

English message

SMS

Type SMS here

Number of characters used: 0

Email

Subject Type Email subject here

Type Email here

Figure 26: Outbound alerting service (beta)

8.3 Development

Development work for the Afrivet services entailed the development of two independent IVR services, one for vets to report incidents and one for dip-tank attendants to report on their activities. A web service was also developed, including the following: 1) an interface for registering dip tank attendants and vets, 2) an interface enabling transcription of user voice

data recorded by the IVR services, and 3c) an outgoing message interface enabling creation of SMS and email messages to be sent out to vets and dip tank attendants.

The web service also made use of the generic content management system, enabling IVR service activity to be monitored at a glance, as well as enabling the generation of usage reports.

Both of the IVR services were developed and rolled out using the existing Lwazi telephony platform software, consisting mainly of the Asterisk (www.asterisk.org) software public branch exchange (PBX) and MobilIVR python application programming interface (API). The web service was developed using the Django web framework (www.djangoproject.com). The outgoing messaging service relied on the use of SMS and email software interfaces included in the Lwazi platform.

The entire Afrivet service was deployed onto external hosting infrastructure, in contrast to previous services rolled out using internal CSIR Meraka infrastructure. More details on this are provided in the platform development section regarding hosted infrastructure (see section 13.3.2).

8.4 Piloting

8.4.1 V-Plan Vet line

The V-Plan Vet line was initially concept tested with 3-5 vets through the Wizard-of-Oz technique, followed by a first pilot of the prototype Vet line (version 0.1) over the period of 9 to 21 July 2012. This consisted of a two-week intensive user engagement period with 14 vets and four internal Afrivet users (with frequent user contact through email and SMS reminders). The usage from this pilot was analysed quantitatively (call log analysis) and qualitatively (user experience interviews). This feedback was used to engage in a consultative redesign process with Afrivet, where revisions were made to both the voice-based service and the web interface. This led to the version 1.0 release of the Vet line service on 2 October 2012.

The version 0.1 Vet line release was accompanied by a brochure which was emailed to the registered users. This aimed to inform them about the line, its purpose, benefits and how and when to use it.

The version 1.0 release was likewise done by means of an email from Afrivet informing the users that the revised line was now in use. This was affirmed in weekly SMS and email reminders sent to the users to request them to report any disease-related matters which they may encounter. At the time of writing this report, the SMS and email reminders were still being sent out weekly, according to the preferences stated by the users at registration.

New requests for registration on the system, as well as error exits were being dealt with by the Meraka HLT team. Afrivet was in the process of taking over the costs for running both the Vet line and the Dip tank line.



Figure 27: Front page of V-Plan Vet line brochure

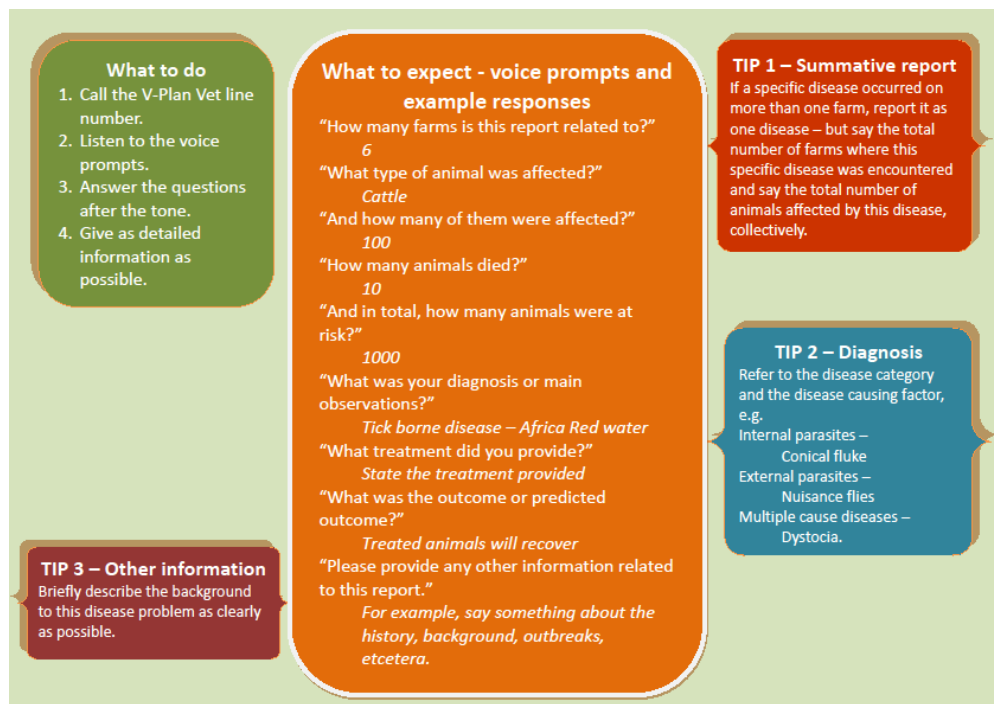


Figure 28: Back page of V-Plan Vet line brochure

8.4.2 Dip tank line

Two versions (0.1 and 1.0) of the Dip tank line were introduced to the dip tank attendants at Siyaphambili which is located in the Ladysmith/Bergville area. In this area there are an estimated 34 dip tanks. Each village in Siyaphambili is allocated a particular day of the week for dipping.

The first site visit (pre-pilot visit) took place on 21 August (introduction of version 0.1) and the second site visit (the piloting of version 1.0), on 18 September 2012.

The version 0.1 pre-pilot visit was specifically aimed at –

- informing the dip tank attendants (users) about the line that was being developed for them to provide feedback on their dipping practice and experiences;
- demonstrating the proposed system to the users;
- ensuring that the system was understood;
- observing the use of the proposed system in order to gauge its usability;
- determining design changes required in consultation with the users; and
- answering any queries that the users had about the proposed system.

The version 1.0 pilot was specifically aimed at –

- introducing the revised (version 1.0) system to the users by means of a demonstration;
- giving the users the number to call after every dipping session;
- providing the users with an opportunity to use the system with an HLT team member present to provide assistance where required;
- answering any queries that the users had about the system.
- encouraging the users to call into the system after each dipping session.



Figure 29: Dip tank attendants attending the pilot

8.4.2.1 Methodology

In order access the Dip tank line, users need to be registered on the system. The contact details of all the dip tank attendants in the area were obtained with the help of Afrivet, and these attendants were then registered as users by the Meraka HLT team, prior to the pilots. The registration was done via a web interface designed for this purpose. Once registered, the system recognises the user through his/her telephone number when he/she calls into the service.

On each of the site visits (pre-piloting and piloting), the dip tank attendants were given a consent form to complete (see annexure F). During the pre-pilot visit (August), the HLT team read the consent form aloud and explained each paragraph in isiZulu as the forms were only available in English. On the piloting day (September), the forms were available in isiZulu and English, and attendants who had not attended the first piloting day, were asked to complete consent forms. As some dip tank attendants arrived late on the piloting day, the Meraka HLT team split up and explained the project and consent form to attendants individually. All dip tank attendants present agreed to participate in the pilot. The team then assisted the dip tank attendants to sign the forms. Questions were answered and clarifications made for those dip tank attendants who had not attend the pre-piloting session. Each dip tank attendant was given a copy of the signed consent form to keep for future reference.

After the consent forms had been collected, the team demonstrated the system, using a speaker-phone. The large size of the group necessitated that the demonstration be carried out three times. A few attendants tried the system using the demo phones supplied by the Meraka HLT team. Some attendants used their own phones or researchers' phones.

The dip tank attendants were not asked formal questions. This would be done at a later stage, once they had used the system for a while. During piloting, an informal approach was followed when demonstrating the system, using the system, and observing the dip tank attendants' use of the system.

On the piloting day, the team explained the difference between versions 0.1 and 1.0 of the system, before demonstrating the system to the attendants and giving them an opportunity to try it out themselves.



Figure 30: Dip tank attendants testing the system

8.4.2.2 Observations (version 1.0 piloting)

The group which attended the pilot session in September was a combination of users who had attended the pre-pilot session and users who had not attended the pre-pilot session. Dip tank attendants who attended the pre-pilot session were more comfortable with the system. They were more likely to volunteer to test the new line. One of the users was not even listening to the prompts; he would enter the responses via DTMF (i.e. touch-tone) input. This had been the method of input for the version 0.1 system but was changed after the pre-pilot session indicated that the users were struggling with the numerical inputs.

One of the younger male attendants who had not attended the pre-pilot session was reluctant to sign the consent form because he did not understand what the project was about as he had arrived late. One female dip tank attendant sitting next to him struggled to explain the system to him. A Meraka HLT team member sat with both of them to explain the project, the Dip tank line and the consent form.

Those who used the Meraka HLT team's phones took long to respond to the prompts because it took a while for them to familiarise themselves with the buttons on the phones.

Problems were encountered with regard to signing the consent forms. The reasons for this was due to a number of the attendants not being able to read the consent form, even though it was in their home language (isiZulu). Some had left their glasses at home, and some have children at home who would read it for them.

Generally, the dip tank attendants were happy with the new line. They volunteered to explain it to their colleagues who could not attend the pilot session. What they were most happy about was the fact that the system would enable the Afrivet team to assist them sooner through the regular reporting, than what was the case with the monthly meeting system which was in place.



Figure 31: Dipping products supplied by DAFF and DEA

8.4.3 Questions

The dip tank attendants asked the following questions about the system:

- What time can we call?
- What if the problem we have is not mentioned?
- Can I use my phone?
- Can I register someone who could not attend?
- What if I do not have airtime?

8.4.4 Challenges

The following challenges were experienced:

- Users took time providing their inputs on the phone. This led to the call being cut off, or, at times, to the system moving to the next prompt, causing confusion.

- Background noise in the hall was a problem, as there were three to four conversations (either calls or explaining the system) taking place at the same time.
- A few older users had to be coached for some time before they were confident enough to call into the system.
- Some users called into the system using the Meraka HLT team's phones with which they were not familiar. This caused confusion, especially when having to make a numerical (i.e. DTMF) input.
- At some dip tanks, there is more than one dip tank attendant and each attendant has his/her own duties to attend to. It may happen, therefore, that a single dip tank attendant may not have all of the information required by the Dip tank line.

8.5 Evaluation

8.5.1 V-Plan vet line

8.5.1.1 Process

The version 0.1 Vet line evaluation took place by means telephonic interviews with 12 of the 14 registered users, over a period of three days (22-24 July 2012).

The information presented here has been categorized according to themes that emerged from the data analysis, but no weighting has been applied.

There were two groups of respondents, namely **users** and **non-users**.

The user group respondents were asked three questions:

- What did you think of the system? (Probe why?)
- Would you recommend the system to your friends/colleagues? (Probe why?)
- Did you know what to do when you called the system? (Probe why/not?)

Non-users were simply asked why they had not used the system and their answers were recorded.

8.5.1.2 Data analysis

In response to the question on their **opinion** of the system, the **user group** indicated that they liked it, and it worked really well. They indicated that the questions were understandable, clear and well-articulated, and that they were afforded enough time to leave a message. They were particularly positive about the prompts at the end of the call flow, which allowed them to leave additional information which they believed might be useful for the person compiling the reports from the IVR input. However, there was an opinion that it took too long to report on one illness at a time. A request was made to allow multiple responses to one prompt and/or multiple reports per animal type. A few respondents indicated that they were not always sure what level of detail was required in

their response to a prompt. They indicated that it was at times difficult to describe the circumstances/situation and the probable outcome.

The respondents reacted positively to the question of whether or not they would **recommend** the service to others, indicating that they believed it to be a quick and easy way of collecting information and relaying this to inform vets on what was happening in their area. They indicated that the information could be enhanced and be made more useful by adding farmers as users. There was a general consensus of opinion that the system would be really useful if it could replace the paper-based reporting system. One respondent indicated that he believed it would benefit the industry if it could enhance the functions of the State.

A few **problems** were encountered early on in the testing. As the piloting was all done by email, users were required to follow the “instructions” in the brochure (Figure 27 and Figure 28). Some called the help line number (set up to help users if they encountered problems using the system) instead of calling the Vet line number, which caused some confusion. One user had difficulty accessing the hash key on his phone’s keypad. Some reported that the transition to the next question was very slow at times. As indicated above (section 8.2.2), these respondents used the DTMF version, and did not realise that they needed to press hash each time after giving their input, in order for the system to move on (they had been prompted to do this only in the first two prompts).

Overall, however, users were a little uncertain the first time they called into the system, but were comfortable using it thereafter. A few respondents indicated that the main challenge would be to ensure uptake of the service.

There were three **non-user respondents**. Their reasons for not using the system included having been on holiday and having been too busy to try it out. The non-users had questions about the system, including whether or not it would replace the paper-based reporting and wanting to know what would happen with the information they submitted. It seems from these responses, that these users could simply be late-adopters, preferring to wait and see what happens with the project, before getting involved.

8.5.1.3 Recommendations

The following recommendations were made:

- The line could also be used for crisis outbreaks of non-notifiable diseases. It need not only be used for reporting notifiable¹⁵ diseases.
- An emergency situation question could be added which, when responded to, would indicate to Afrivet that there was an issue which required immediate attention.

¹⁵ Vets are required by law to report certain diseases to the State veterinarians.

- Reporting should be required on a weekly basis not daily (as had been requested during the piloting). In fact, the general feeling was that the line should only be used when one really had something important to report.
- There was a suggestion that a reminder service be implemented for vets to follow up on the cases reported on in previous weeks, if required.
- SMS reminders to call in and report on diseases encountered should be sent to vets. (This was in fact part of the initial design.)

8.5.2 Dip tank line

Structured key informant interviews (using an interview questionnaire) were conducted with dip tank attendants who consented to participate in the evaluation, in order to obtain feedback on their experience of the Dip tank line (version 1.0) and gauge the usability of the system. The information obtained during the evaluation will be used to improve future versions of the line.

8.5.2.1 Process

A group of 21 dip tank attendants was invited to meet with the research team on 20 November 2012.

On arrival the team introduced themselves. The consent forms in isiZulu were handed out to verify participation in the evaluation. The form was read out in isiZulu and all the different sections were explained. There were a few dip tank attendants present who had not attended the pilot, and as a result extra time was spent explaining the project to them. A total of 19 dip tank attendants signed the consent form and agreed to participate in the evaluations.

The questionnaire was read out aloud to the whole group. Questions and misunderstandings were addressed. Individual interviews were then held with the respondents. Each interview lasted about 15 minutes. The interviews were conducted in isiZulu. If it was established that a respondent had not used the system, the interview was terminated.

8.5.2.2 Results

Dip tank attendants responded on their experience of the system as follows:

- There was an increased awareness and importance of dipping.
- Others had learnt about the line.
- The line forced dip tank attendants to count their stock and in a way that is more organized.
- They liked the fact that the service called them.
- It saved money on travelling to the vet.

- The system was free.
- They were able to get feedback quickly.
- It was not easy at first but most said they found it easy to use, including following prompts and responding to them.
- It was easy to use both buttons and to say one's response.
- It allowed them to report problems immediately and was easy to use.
- It provided the ability to report uncommon diseases, provided knowledge and enabled immediate solutions to problems.
- It enabled them to get help when animal technicians were not available.
- They did not have to wait for the monthly meetings.

8.5.2.3 Challenges

- Many respondents were not aware that they could “barge in” (respond while the prompt was still playing), and tended to wait and listen to the whole question before responding.
- Other users wanted to speak while one of them was busy on the line.
- There were some who could not call in because there was no clear network or they had lost their phones.

8.5.2.4 Recommendations

- The system needs to be shared with others in the same dip tank area in case the person who has been taught is unable to dip.
- The system must ask questions about water shortages.
- The system should be able to provide advice at the end, for example, to remove thorn bushes.
- The system should provide an opportunity to say everything at once, instead of asking one question at a time.
- CSIR should provide funding to the livestock owners.

8.5.3 Conclusions

The Dip tank line was found to be useful by the targeted users. It has a number of benefits including saving money, helping to ensure that they count their stock, getting immediate assistance etc. The users recommended that the line be rolled out to more users.

8.6 Call log analysis

This section of the report presents the call log analysis of calls placed to the Afrivet Dip tank line and Vet line. This call log analysis presents a quantitative view towards understanding the usage of these services. Various metrics such as total number of calls, call duration, call frequency, number of reports logged, and also call patterns such as final states, invalid

entries and timeouts that occur in a call are analysed to obtain a thorough understanding of the users' interactions with the services.

8.6.1 Dip tank line

The call log analysis showcased in this section covers the calls made to the service by dip tank attendants over the period of September to December 2012. The Dip tank line was piloted on 18 September 2012 in Ladysmith, KwaZulu-Natal, and at the time of this analysis the service¹⁶ had been running for 12.3 weeks at the pilot site. The call log analysis was conducted only on calls made by the dip tank attendants after the piloting day, which ensures the calls being represented herein were not calls made during the demonstration of the system to the dip tank attendants. The Dip tank line is provided in two languages; isiZulu and English, but all the dip tank attendants preferred to use it in isiZulu.

8.6.1.1 Call volume

Dip tank attendants called the system 106 times from September to December 2012. Figure 32 below, shows the time distribution of the calls received by the service over this period. We notice that most calls were made between October and November, with a few calls being made in September, after the pilot. The higher usage in October and November can be attributed to the fact that a local Afrivet representative actively and regularly engaged with the dip tank attendants on service usage. User experience interviews were also conducted by the research team during this time.

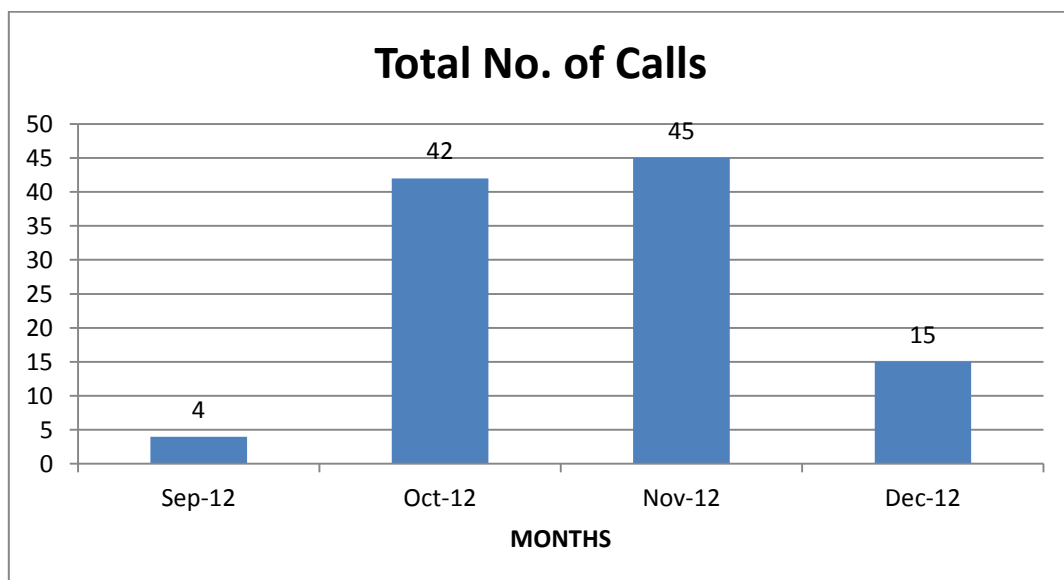


Figure 32: Total call distribution per month

¹⁶ 18 September - 13 December 2012

In Table 27 we further analyse this data to calculate the average calls per week which proportions the total calls received per month with respect to the number of weeks the pilot has been running. We observe that there were at least 3 calls per week during over the October to November period.

Table 27: Average calls per week

	Months			
	Sep-12	Oct-12	Nov-12	Dec-12
Calls	4	42	45	15
No. of pilot weeks	12.3	12.3	12.3	12.3
Calls/week	0.3	3.4	3.7	1.2

8.6.1.2 Call duration

Average call duration was found to be 2 min and 21 s. Figure 33 illustrates the distribution of the call duration. We observe that 10% of the calls range from 15 s to 1 min, a significant 86% of the calls range between 1 min 15s and 3 min 15 s, and 4% of the calls end at 3 min 30 s or more.

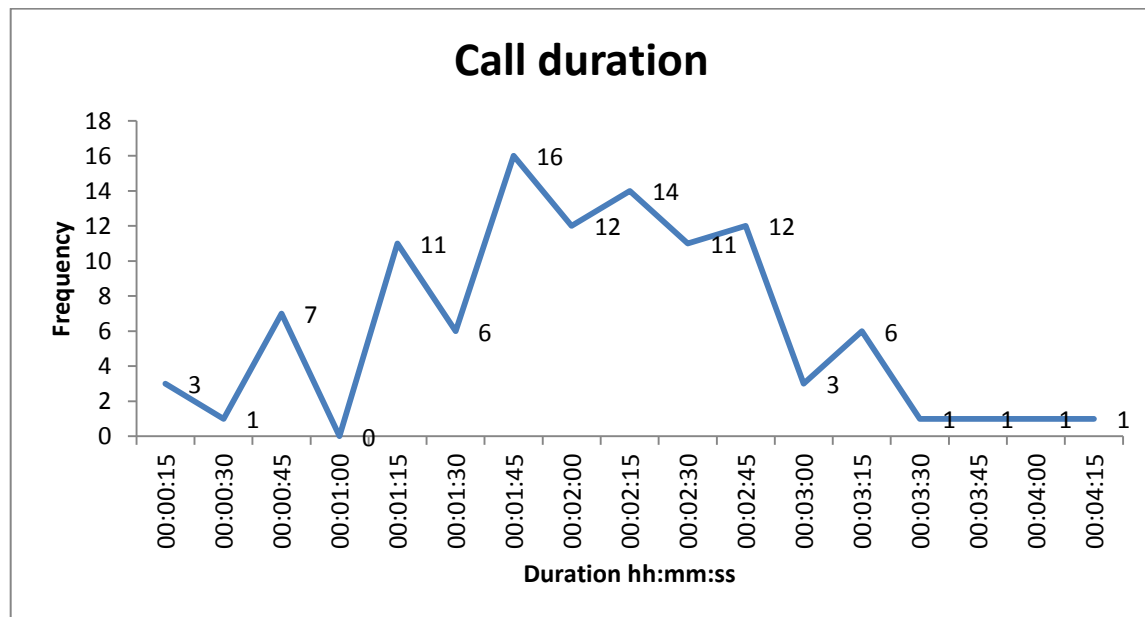


Figure 33: Call duration distribution

8.6.1.3 Usage frequency analysis

Usage frequency analysis presents the number of times a particular user calls the service across the analysis period. Across the period of September to December 2012, the service received a total of 106 calls. We observe that many of the dip tank attendants called the service at least 3 times or more per month (see Table 28). At the time of analysis, there were

31 dip tank attendants registered on the service; at least 20 of them had called the system, with about 50% of them having called the line around 5 times over the piloting period.¹⁷

Table 28: Usage frequency distribution

Phone numbers	Dates				Grand Total	%
	Sep-12	Oct-12	Nov-12	Dec-12		
728262044	-	12	5	3	20	18.90%
731605505	-	5	5	3	13	12.30%
761189151	-	4	5	1	10	9.40%
765202754	-	3	5	2	10	9.40%
730404204	2	5	1	-	8	7.50%
785915879	-	5	3	-	8	7.50%
727561106		3	3	-	6	5.70%
796158426	2	1	3	-	6	5.70%
793686597	-	1	1	3	5	4.70%
717080458	-	-	4	1	5	4.70%
712140150	-	-	3	1	4	3.80%
724576881	-	-	3	-	3	2.80%
792276915	-	-	3	-	3	2.80%
732609726	-	2	1		3	2.80%
763594285	-	-	-	1	1	0.90%
792226586	-	1	-	-	1	0.90%

8.6.1.4 Final call state

The final call state represents the last dialog state (or step) in the call flow of the service where the user ended the call. It gives an indication whether the users were able to complete the task of providing feedback about their dipping session, any diseases the animals might have, or any other problems related to dipping. It also provides more insight into the possible areas of the call flow where users are dropping off, which might require further attention. Figure 34 illustrates the final call state for calls received across all the calls made by dip tank attendants.

¹⁷ The frequency of calls expected from the dip tank attendants varies based on their dipping frequency and season. Over the piloting period, this would have ranged from weekly/bi-weekly (summer) dipping to only dipping every three weeks (towards September 2012 – slightly cooler).

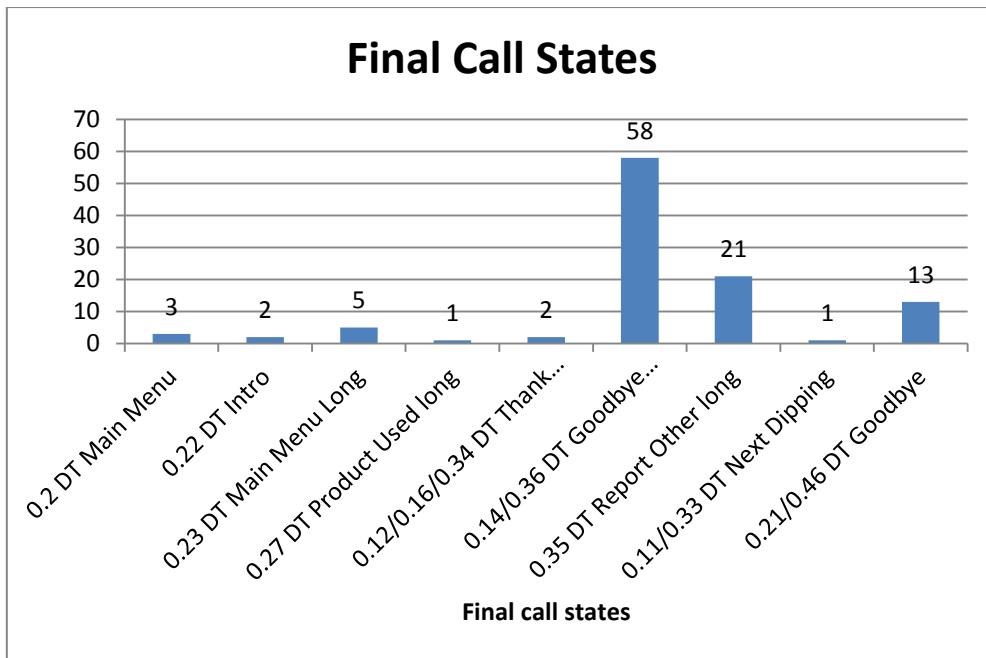


Figure 34: Final call states

Table 29 shows a breakdown of the final call states across all the calls. We find that 9% of the calls end at the first few dialog states of the call (0.2 DT Main Menu, 0.22 DT Intro & 0.23 DT Main Menu Long), in another 2% of the calls the user hung up mid-way through the call and hung up whilst still answering questions about the dipping (0.27 DT Product Used long and 0.11/0.33 DT Next Dipping).

The majority of calls (77%) ended successfully, where the user had gone through all the dipping-related questions until the last state (0.12/0.46 DT Goodbye, 0.14/0.36 DT Goodbye share, 0.35 DT Report Other long) where he/she had been thanked for calling the service. The remaining 12% of the calls ended at an “error exit”, where the user had provided invalid input or no input three times consecutively and was then asked to leave a message for further assistance on how to use the service.

Table 29: Distribution of final call states

Final State	Total	%
0.2 DT Main Menu	3	9%
0.22 DT Intro	2	
0.23 DT Main Menu Long	5	
0.27 DT Product Used long	1	1%
0.11/0.33 DT Next Dipping	1	1%
0.12/0.16/0.34 DT Thank you	2	77%
0.14/0.36 DT Goodbye share	58	

Final State	Total	%
0.35 DT Report Other long	21	
0.21/0.46 DT Goodbye (error exit)	13	12%
Grand total	106	

8.6.1.5 Invalid entries

Invalid entries (or wrong input) refer to cases where the user pressed an invalid key when prompted by the service to choose from a defined set of keys at a particular dialog state. For example, some questions in the service provide the user with the option to press 1 or 2; if the user presses any other key besides 1 or 2 at such a question, this is logged as an invalid entry. When the user provides a wrong input for the first and second time, he/she is re-prompted by the service appropriately to answer the question with correct input only. The third time, the service requests the user to leave his/her name and number for a representative to call him/her back should he/she require further help, and then exits the call (this is called an “error exit”). Table 30 illustrates the overall wrong inputs experienced across the service. We find that there were only three calls (2.8% of total calls) which had a single invalid entry per call.

Table 30: Invalid entries

No. of invalid entries	No. of calls
1	3
Grand Total	3

8.6.1.6 Timeouts

A timeout refers to a case where a user provides no input (does not press any key) when the service asks a question. Similar to the invalid entry (wrong input), the service will re-prompt the user twice for input using appropriately designed prompts, and third time the service will ask him/her to leave his/her name and number for a representative to call him/her back, and exit the call (an “error exit”).

Table 31 provides an overview of the total number of calls in relation to the frequency of timeouts. We find that 15 calls (14% of the total number of calls) had more than one timeout, i.e. in 15 calls the user did not provide any input at least once, 4.7% of the calls (5 calls) had two timeouts and three calls (11%) had three timeouts.

We observe that even though users did have timeouts, they were still keen to continue with the call when re-prompted and to proceed with the call by providing some kind of an input (as evidenced by the fact that there are 15 and five calls with one and two timeouts respectively, which means that the user recovered from the timeout, versus the calls with three timeouts where the service has to exit the call).

Table 31: Timeouts

No. of timeouts	No. of calls
1	15
2	5
3	12
Grand Total	32

Looking at the number of invalid entries (a total of 3) versus the number of timeouts (a total of 32), we see that there is a larger number of the latter, i.e. users choose to provide no input to the service in many more instances rather than providing invalid (wrong) input.

8.6.1.7 Barge-in

Barge-in refers to when a user consciously interrupts a menu prompt to make his/her choice by pressing a key (i.e. does not wait for the prompt to finish). This usually implies that the user is comfortable with the menu options and knows which option to choose, and signifies that the user is becoming familiar with the service.

From Table 32, we observe that 61% of calls have no barge-in (i.e. the user did not interrupt the menu prompt and waited until the prompt had finished saying all the options, before pressing a key to indicate their choice). We see that 24% of the calls had one barge-in (i.e. the user interrupted the menu prompt at least once in the call), 13% had two barge-ins, and only 2% of the calls had three barge-ins, i.e. an overall barge-in rate of 39% (24+13+2%) across all calls.

Table 32: Barge-in distribution

No. of barge-ins	Total	%
0	65	61%
1	25	24%
2	14	13%
3	2	2%
Grand Total	106	

8.6.2 V-plan vet line

The V-Plan Vet line (version 0.1) was initially piloted with a subset of users for limited period of time in order to assess the usability and user experience of the service. The first pilot was conducted with 14 vets and four Afrivet users for a period of 10 days. The service was then redesigned extensively based on the call log analysis and the qualitative telephonic

interviews conducted to gather user feedback. The second version of the service (version 1.0) was launched in October and was still running when this analysis was made.

8.6.2.1 First pilot: Call log analysis

This call log analysis covers the period 9 to 21 July 2012, when the first pilot of the Vet line (version 0.1) ran for approximately 10 days (excluding weekends).

8.6.2.1.1 Call volume and reporting frequency

Users made **47 calls** to the service over the period of 9 to 21 July 2012. Figure 35 below, shows the time distribution of the calls received by the service over this period; these calls were made by both private vets and Afrivet internal users.

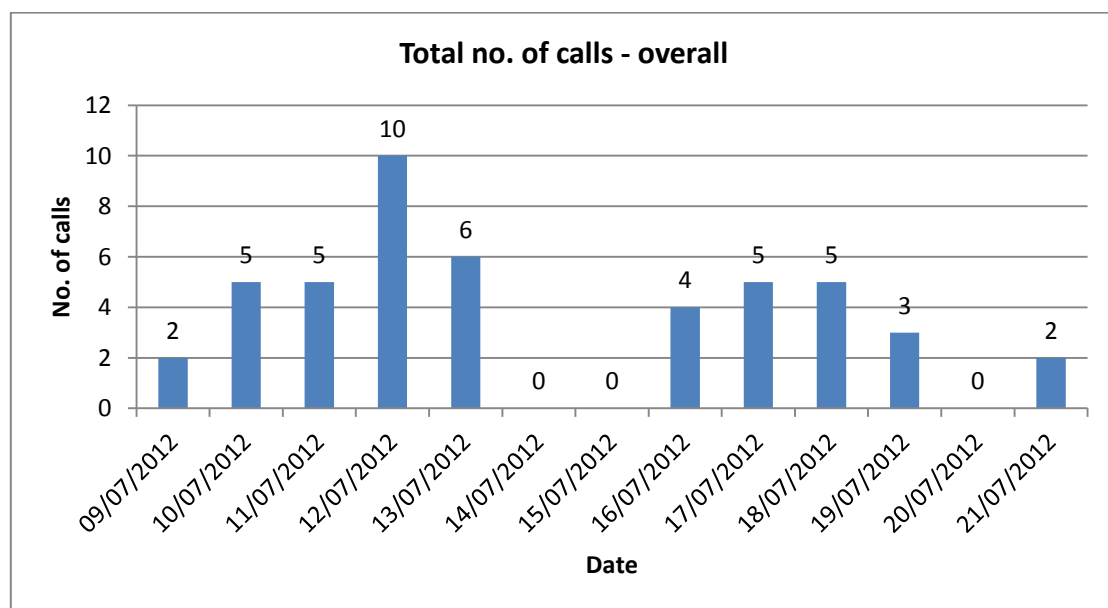


Figure 35: Total call distribution per day – Overall

Figure 36 shows the total number of calls made by private vets only. In both graphs (Figures 4 and 5) we notice that the maximum number of calls on one day was nine calls, made on 12 July 2012. This can be attributed to the active engagement by an Afrivet representative with the private vets.

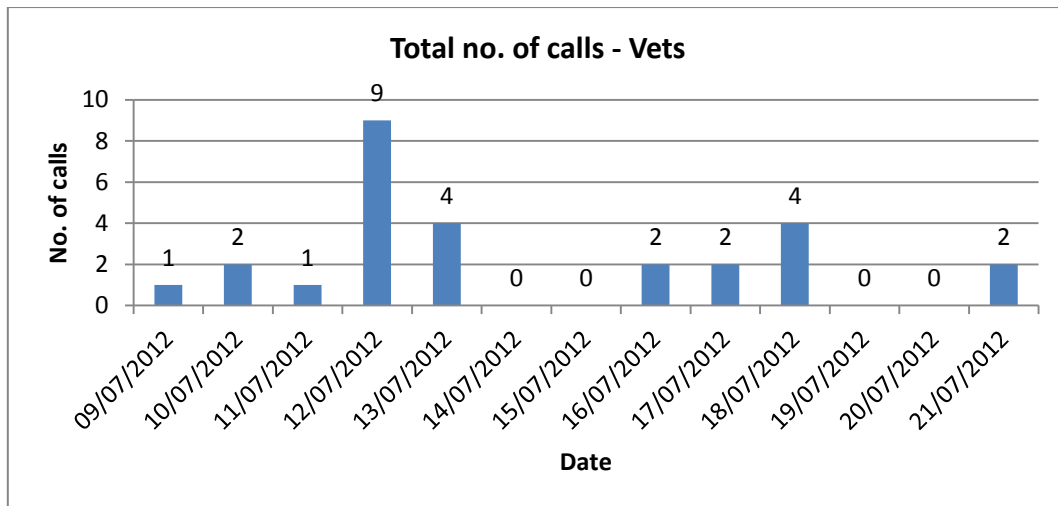


Figure 36: Total call distribution per day – Vets

From Table 33 we note that the service had 14 registered private vets and four Afrivet internal users; constituting a total of **18 users** registered on the service over this period. We also note that the private vets made 27 calls to the system, while the Afrivet users made 20 calls.

Table 33: Registered users and calls received

Registered callers	No. of registered callers	Calls received	Average
Private vets	14	27	2.7
Afrivet	4	20	2
Total	18	47	4.7

We further analysed how many calls were successful, unsuccessful, or had multiple reports (see Table 34). Successful calls are calls in which users have gone through all the states of the call flow (prompts) and have finished the task successfully. Unsuccessful calls are calls in which users did not go through all the states of the call flow successfully, i.e. the call was dropped at a particular state. Multiple reports refer to calls that have more than one report, for example, in cases where users submitted a 2nd or 3rd disease incidence in one call. From Table 34 we observe that 89% of the calls (24+18 calls) were successful, 11% of the calls (3+2 calls) were unsuccessful and only four calls had multiple reports.

Table 34: Call success rate

Users	Successful	Unsuccessful	Multiple reports
Private vets	24	3	2
Afrivet	18	2	2
Total	42	5	4

8.6.2.1.2 Usage frequency analysis

Usage frequency analysis presents the number of times a particular user calls the service across the analysis period. Across the analysis period the service received 27 calls from vets. The caller frequency is indicated in Table 35 below, where we find that of the 14 registered users, 11 users called the service at least once, while there were three non-users.

Table 35: Usage frequency distribution – Vets

Vet name	09-Jul	10-Jul	11-Jul	12-Jul	13-Jul	14-Jul	15-Jul	16-Jul	17-Jul	18-Jul	19-Jul	21-Jul	Total calls
F van Niekerk				1	1			1					3
D Lessing				3								2	5
D van Zyl										1			1
P Nel	1		1		1								3
L van Jaarsveld					1			1		1			3
S Wessels				1						1			2
C Nel									1				1
D Hauptfleisch				1									1
N Degenaar				1									1
M Smith		2		1					1	1			5
H Pieters				1	1								2

We also analysed the time the vets would typically call the service, to obtain insight into appropriate times for sending automated reminders for reporting, since the vets are on the road most of the time. Figure 37 illustrates this analysis, where we see that most vets called between 12:00 and 18:00, or later.

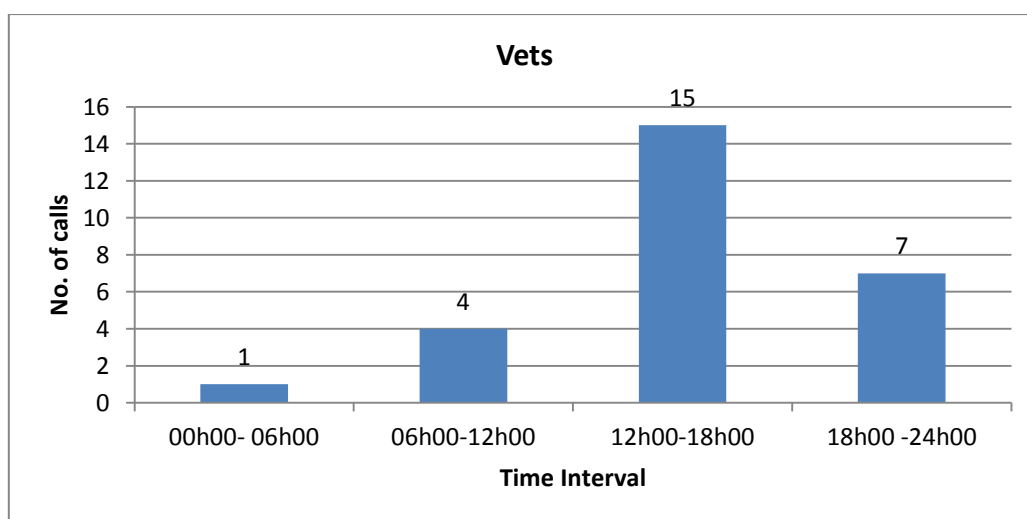


Figure 37: Call time distribution – Vets

8.6.2.1.3 Call duration

We analysed call duration for calls made by users who reported once (single report), as well as those who made multiple reports (on successful calls only). Table 36 & Table 37 illustrate the distribution of call duration for both reports. We observe that the average call duration for single report was found to be 2 min 25 s (shown in Table 36) and average call duration for multiple reports was found to be 4 min 36 s (shown in Table 37). We also observe that the average duration of the audio message left by the users on the service was 58 s.

Table 36: Call duration – Single reports

Single report	Duration (hh:mm:ss)
Average	00:02:25
Median	00:02:19
Min	00:01:37
Max	00:05:17

Table 37: Call duration – Multiple reports

Multiple reports (2)	Duration (hh:mm:ss)
Average	00:04:36
4 Calls with durations: 00:03:24, 00:04:50, 00:04:52, 00:05:17	
Min	00:03:24
Max	00:05:17

8.6.2.1.4 Other call patterns analysed

We further analysed other call patterns such as timeouts, invalid entries and error exits. A timeout occurs when a user provides no input (does not press any key) when the service asks a question. We note that there were five timeouts across all the calls (two for private vets, three for Afrivet internal users).

Invalid entries (or wrong input) refer to a case where the user presses an invalid key when prompted by the service to choose from a defined set of keys at a particular menu. We found that there were no invalid entries in any of the calls.

An error exit occurs when a user makes three consecutive timeouts or invalid entries, and the service asks him/her to leave a message for further assistance and exits the call. We only found one error exit in these calls.

8.6.2.2 Second launch: Call log analysis

Based on the first pilot analysis and user feedback, the Vet line was extensively redesigned to further streamline the service and then launched on 2 October 2012. At the time of this analysis¹⁸ the service had been running for 10.3 weeks. The service had 17 registered vets (excluding Afrivet’s internal users) and was provided in English and Afrikaans, with 15 users using Afrikaans and two users preferring English.

8.6.2.3 Call volume and reporting frequency

We found that the vets had made nine calls to the Vet line over the period October to December 2012. Table 38 below, shows the time distribution of the calls received by the service from different vets during this period; of these, seven calls were made in October, only two calls were made in November, and there were no calls in December.

Table 38: Total call distribution per month

Months	Vet practice name				Total
	Frankfort Dierekliniek	Humansdorp Veterinary Clinic	Onderstepoort Veterinary Academic Hospital	Wilgepoort Dierekliniek	
10/2012	1	4	0	2	7
11/2012	0	0	2	0	2
Total	1	4	2	2	9

Table 38 also highlights the reporting frequency which provides information on the number of times a particular vet (or vet clinic) used the service to report disease incidences. We observe that the highest number of reports submitted was four (out of a total of nine calls), and that these were placed by the Humansdorp Veterinary Clinic in October 2012.

The Vet line also allows vets to leave multiple disease incidence reports within a single call. We analysed this feature’s usage and found that in eight of the nine calls, vets left two disease incidence reports per call (in the remaining one call a single disease incidence was reported). In Table 39 we further analyse this data to calculate the average calls per week, which proportions the total calls received per month with respect to the number of weeks the pilot was running when the analysis was made. Here we see that the vets barely called the service once a week.

⁵ 2 October – 13 December 2012

Table 39: Average calls per week

Months	Oct-12	Nov-12
Calls	7	2
No. of pilot weeks	10.3	10.3
Calls per week	0.7	0.2

8.6.2.3.1 Call duration

Average call duration was found to be 1 min and 41 s. Figure 38 illustrates the call duration distribution, where it can be seen that most of the calls lasted between 1 min 15 s and 2 min 30 s.

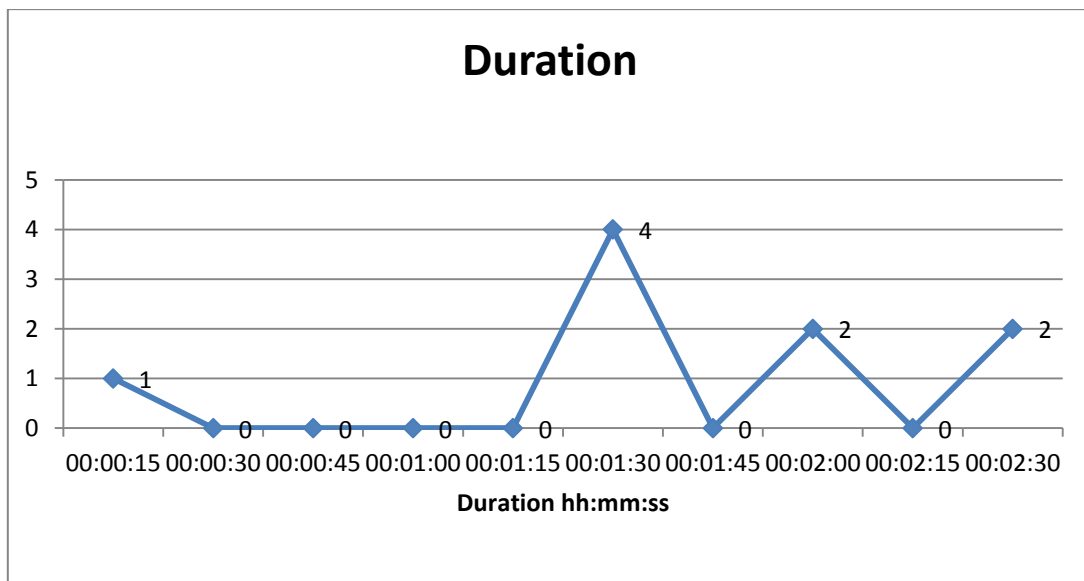


Figure 38: Call duration distribution

8.6.2.3.2 Final call state

The final call state represents the last dialog state (or step) in the call flow where the user ended the call. It gives an indication of whether or not the vets were able to successfully complete the task of providing a disease incidence report. Figure 39 illustrates the final states for all the Vet line calls over this period.

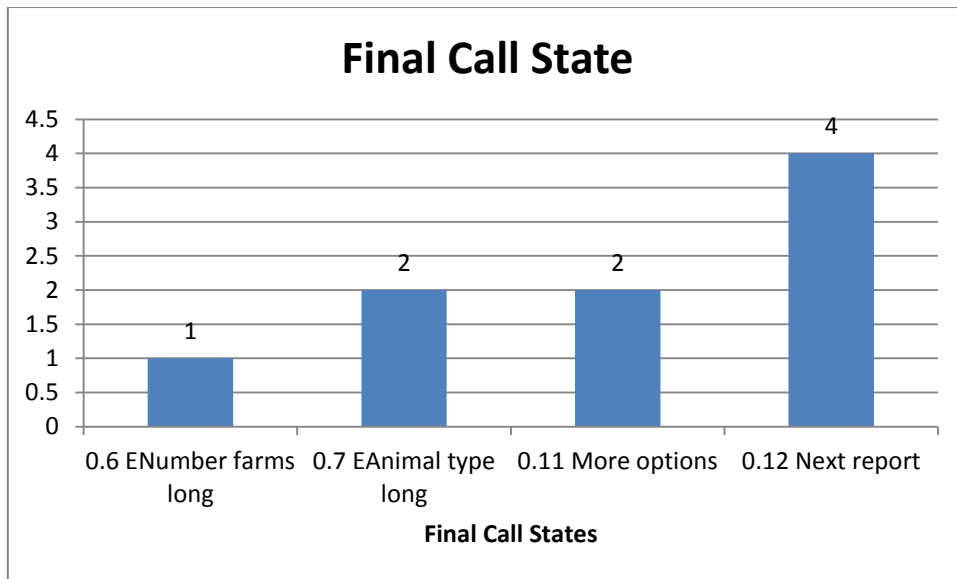


Figure 39: Final call states

Table 40 shows that, proportionally, four out of nine calls end at state “0.12 Next report” (where users can continue with the next report), which means the user successfully completed one report. Two calls end at “0.11 More options” where users have the option of leaving another report or hanging up. The other three calls end at call states “0.6 ENumber farms long” and “0.7 Animal type long”, where users were still being asked questions for the first disease incidence report (this implies an incomplete report).

Table 40: Distribution of final call states

Final States	Total	%
0.6 ENumber farms long	1	11%
0.7 EAnimal type long	2	22%
0.11 More options	2	22%
0.12 Next report	4	44%
Grand Total	9	

8.6.3 Discussion

Based on the call log analysis for the Dip tank line and the V-Plan Vet line in the preceding sections, we observe that the Dip tank line has a much higher usage than the Vet line.

On average, across the analysis period, there were three or more calls per week to the Dip tank line, with at least 66% of the registered dip tank attendants being active on the service. Very few (3%) invalid entries (wrong inputs) were experienced by the users, with close to 30% of the calls experiencing one or more timeouts. Of these, 19% were able to recover from the timeout and proceed with the call, and 11% of the calls were directed to an “error exit” with the user being asked to leave a message to request further assistance. We observe

that close to 61% of calls had no barge-in (i.e. the user did not interrupt the menu prompt, and waited until the prompt was finished before pressing a key to indicate their choice) and 39% had barge-in. The higher lack of barge-in can possibly be attributed to the fact that many of the dip tank attendants were still getting used to the service and preferred to listen to all of the options before making a menu choice. However, the 39% of users using barge-in is encouraging as it indicates that these users felt comfortable enough to interrupt the menu prompt and move on with the call.

One of the crucial factors which contributed to the more consistent usage of the dip tank line is the active engagement of an Afrivet representative with the dip tank attendants to regularly follow-up on their calls (or lack thereof). We found that this active engagement builds trust in the service and showcases to the dip tank attendants that the information they provide is actually being used to assist them with their dipping activities, i.e. closing the feedback loop.

The V-Plan Vet line has seen significantly low uptake; we believe that this may be due to the vets' existing mental model around using the service as a reactive emergency disease reporting line, as opposed to a proactive line for daily reporting of disease incidences seen in the area, for trends mapping. We conjecture, given the correct marketing and incentives with the vet user group, as well as expanding the user base to livestock owners in the community as well as state vets, the Vet line could play a critical role in disease surveillance in the country.

9. Services-related HCD and knowledge generation

Although there were no post-graduate studies relating to the Lwazi II services only, several publications emanated from the services subproject (see section 16.3.4), and three candidate researchers (Olwethu Qwabe, Tshepo Moganedi and Mpho Kgampe) and one intern (Akhona Samson), obtained in-service training on various aspects of speech application implementation, including selection, requirements analysis, design, software development, piloting, interviewing, data analysis and report-writing.

10. Services conclusions

Table 41 summarises the pilot sites, languages and domains relating to the services piloted and deployed during Lwazi II. All services-related project requirements were successfully met.

Table 41: Consolidated view of the Lwazi II services, piloting and evaluation

Service no.	Service name	Pilot no.	Pilot site	Language(s)	Domain	Pilot date	Evaluation date
1	CDW	1	Bushbuckridge (LP)	Xitsonga	Social services	11 May 2010	None (Evaluated in Lwazi I)
1	CDW	1	Musina (LP)	Tshivenda	Social services	2 June 2010	None (Evaluated in Lwazi I)
2	CDW Enhanced	2	Sterkspruit (EC)	isiXhosa, English	Social services	23 & 24 November 2010	None
3	DBE SMS report reminder	3	DBE head office (GP)	English	Corporate Governance	17 January 2011 22 September 2011	Ongoing discussions
4	DBE learner feedback	4	Kokotla JSS (GP)	English, Sepedi, isiZulu	Education	27 July 2011	11 February 2012 17 March 2012
5	DBE coordinator reporting	5	Kokotla JSS (GP)	English	Education	27 July 2011	30 August 2011
4	DBE learner feedback	4	Ntsako JSS (GP)	English, Sepedi, isiZulu	Education	29 July 2011	18 February 2012 10 March 2012 24 March 2012
5	DBE coordinator reporting	5	Ntsako JSS (GP)	English	Education	29 July 2011	30 August 2011
4	DBE learner feedback	6	Inkonjane MS (NW)	English, Sepedi, Setswana, isiZulu, Afrikaans	Education	2 November 2011	27 May 2012

Service no.	Service name	Pilot no.	Pilot site	Language(s)	Domain	Pilot date	Evaluation date
5	DBE coordinator reporting	7	Inkonjane MS (NW)	English	Education	2 November 2011	9 November 2012
4	DBE learner feedback	6	Thuto Pele JSS (NW)	English, Sepedi, Setswana, isiZulu, Afrikaans	Education	3 November 2011	26 May 2012
5	DBE coordinator reporting	7	Thuto Pele JSS (NW)	English	Education	3 November 2011	9 November 2012
4	DBE learner feedback	6	Nqobangolwazi JSS (MP)	English, Sepedi, Setswana, isiZulu, Afrikaans	Education	7 November 2011	2 May 2012
5	DBE coordinator reporting	7	Nqobangolwazi JSS (MP)	English	Education	7 November 2011	None ¹⁹
4	DBE learner feedback	6	Reggie Masuku HS (MP)	English, Sepedi, Setswana, isiZulu, Afrikaans	Education	2 February 2012	13 May 2012
5	DBE coordinator reporting	7	Reggie Masuku HS (MP)	English	Education	2 February 2012	13 May 2012
6	Job search	8	Focus group (GP)	English, Afrikaans	Employment	20 December 2011	20 December 2011
8	Afrivet V-plan Vet line	9	Nation wide	English, Afrikaans	Agriculture	9-21 July 2012 (0.1) 2 October 2012 (1.0)	22-24 July 2012
9	Afrivet Dip tank line	10	Siyaphambili, KZN	English, isiZulu	Agriculture	18 September 2012	20 November 2012

¹⁹ Coordinator was hospitalised and could not be interviewed.

Service no.	Service name	Pilot no.	Pilot site	Language(s)	Domain	Pilot date	Evaluation date
4	DBE learner feedback	11	Mabusa Besala SS (MP)	isiNdebele, isiZulu, English, Sepedi, Setswana	Education	23 August 2012	None ²⁰
5	DBE coordinator reporting	12	Mabusa Besala SS (MP)	English	Education	23 August 2012	None
4	DBE learner feedback	11	Mthombo SS (MP)	siSwati, English, isiNdebele, Xitsonga, isiZulu	Education	16 August 2012	18 October 2012
5	DBE coordinator reporting	12	Mthombo SS (MP)	English	Education	16 August 2012	19 October 2012
4	DBE learner feedback	11	Tshiawela SS (LP)	Tshivenda, English, Sepedi, Xitsonga, Setswana	Education	5 September 2012	None
5	DBE coordinator reporting	12	Tshiawela SS (LP)	English	Education	5 September 2012	None
4	DBE learner feedback	11	Marimane SS (LP)	Xitsonga, English, Sepedi, Tshivenda, isiZulu	Education	4 September 2012	7 November 2012
5	DBE coordinator reporting	12	Marimane SS (LP)	English	Education	4 September 2012	7 November 2012

²⁰ Only 3 schools were evaluated for phase 3.

Service no.	Service name	Pilot no.	Pilot site	Language(s)	Domain	Pilot date	Evaluation date
4	DBE learner feedback	11	Cedar SS (FS)	isiZulu, isiXhosa, Sesotho, English, Afrikaans	Education	27 August 2012	None
5	DBE coordinator reporting	12	Cedar SS (FS)	English	Education	27 August 2012	None
4	DBE learner feedback	11	Ikeketseng SS (FS)	isiZulu, isiXhosa, Sesotho, English, Afrikaans	Education	7 September 2012	None
5	DBE coordinator reporting	12	Ikeketseng SS (FS)	English	Education	7 September 2012	None
4	DBE learner feedback	11	Kwezilomso SS (EC)	isiXhosa, English, isiZulu, Afrikaans, Sesotho	Education	30 August 2012	None
5	DBE coordinator reporting	12	Kwezilomso SS (EC)	English	Education	30 August 2012	None
4	DBE learner feedback	11	Cowen SS (EC)	isiXhosa, English, isiZulu, Afrikaans, Sesotho	Education	29 August 2012	31 October 2012
5	DBE coordinator reporting	12	Cowen SS (EC)	English	Education	29 August 2012	31 October 2012

Subproject – Technologies

11. Text-to-speech (TTS) conversion

11.1 Introduction

For the text-to-speech (TTS) component of Lwazi II, the goal was to develop systems that are demonstrably more natural than systems developed during Lwazi I.

For improvements in synthesis quality and naturalness during this project we focussed on the following approaches:

- Speech synthesis research and development based on Hidden Markov Models (HMMs) rather than unit-selection synthesis, considering the observation that HMM-based synthesis results in more predictable synthesis quality especially when source speech data is limited (Van Niekerk *et al.*, 2009). This included work on acoustic modelling as well as signal processing for synthesis using this technology.
- Collection of larger speech corpora with more prosodic variation (compared to the smaller Lwazi I corpora of neutral prosody).
- The implementation of more advanced natural language processing (NLP) modules in the TTS text-analysis front-end, including text tokenisation, normalisation, and specialised language-specific modules (such as tone in the Sotho languages).

In the following sections we will present activities, results and a summary of outputs generated during the course of the project with links to locations where these can be found/downloaded.

11.2 Activities

This section of the report will be structured into three subsections corresponding to the approaches presented above.

11.2.1 Research and development for HMM-based speech synthesis

Relatively independent streams of research and development work supporting improved TTS synthesis of Lwazi II voices from HMMs are described in the following subsections.

11.2.1.1 Acoustic modelling and synthesis

During the first half of 2010, development efforts focussed on supporting the HMM-based speech synthesizer (HTS: <http://hts-engine.sourceforge.net/>). This involved writing software modules (plugins) for our TTS platform (Speect: <http://speect.sourceforge.net>) and the extension of a number of existing modules, including the accompanying suite of voice creation tools. The system was demonstrated through an entry into the international Blizzard Challenge 2010 (http://www.synsig.org/index.php/Blizzard_Challenge) (Louw *et al.*, 2010) and the software was released as Speect 0.9.5 with Voicetools 0.1.0 on 4 June 2010.

This release also included numerous bug fixes and an enhanced Python interface and documentation. During this effort, a number of new NLP components were also implemented for English (see section 11.2.2). Further development involved rebuilding all Lwazi I voices using the updated Speect platform and HTS synthesizer for the purpose of comparison with new voices developed for Lwazi II (see section 11.3). This version of the TTS platform was also tested through an effort to port a kiSwahili voice from Festival to Speect; this was completed successfully.

During 2011 work was done towards improving HMM acoustic models based on the interim Lwazi II speech corpora. This included the implementation of a spectral distance measure (Kominek *et al.*, 2008) in order to validate the quality of acoustic models, and research into using such a distance measure to automatically detect inaccuracies in recorded TTS corpora. These developments, which are applicable to all speech corpora collected during Lwazi II, were applied to the Afrikaans speech corpus in order to improve acoustic models and are described in Van Niekerk (2011). In addition to this, acoustic modelling using current speaker adaptive techniques (Yamagashi, *et al.*, 2009) was investigated using Afrikaans and English TTS and ASR corpora.

During 2012 acoustic modelling tools were updated to include the modelling of bandpass voicing strengths using HMMs. This formed part of the work to improve synthesis quality by implementing the mixed excitation synthesis algorithm proposed in Yoshimura *et al.* (2001). This was done by extending the HTS synthesizer (<http://hts-engine.sourceforge.net/>) and incorporating this (including updated acoustic modelling tools) into our TTS platform. All Lwazi II TTS systems were updated to make use of this technology.

11.2.1.2 Pitch modelling

Over the course of the project the modelling of pitch, specifically with regard to tone languages was studied (Barnard and Zerbian, 2010; Van Niekerk and Barnard, 2010; Van Niekerk and Barnard, 2012; Van Niekerk and Barnard, to appear). This led to the development of improved software tools for robust automatic pitch extraction from speech (based on the Praat software package (Boersma, 2001)) which were incorporated into Lwazi II development tools. This work also complemented the work on tone described in section 11.2.2.3, with the eventual goal of appropriate speech synthesis of African tone languages, including some of the South African languages.

11.2.1.3 Code-switching

In 2011, prototype systems capable of synthesising English words embedded in primary language text (e.g. Zulu text containing English place names) was built for a number of languages. This was done by developing English acoustic models from the sets of English recordings that were made for some languages (see section 11.2.1.4) and additional modules in the TTS platform allowing the use of English text analysis modules for voices in other languages. This development should extend the applicability of Lwazi II TTS voices,

especially for rendering real-world numerical, date-related and news texts, where primary language and English terms are often interleaved (Grover *et al.*, 2009).

11.2.1.4 Speech corpus development

As described in the introduction above (section 11.1), one of the approaches taken to improving synthesis quality was to record larger speech corpora with more natural prosodic variation. We completed recordings of a set number of languages per year as set out in the project proposal, prioritising recordings of languages required by Lwazi II application pilots and updated the corpus development protocol during the course of the project as described below.

The corpus development effort for each language can be summarised into the following steps:

1. Obtain relatively clean text from a large reference corpus (largely based on various government information documents) by identifying sentences containing valid words in the language according to existing spelling checkers. This was done by the North-West University at Potchefstroom.
2. Select a subset of sentences to include all acoustic units required, by automatic phonetisation and a greedy algorithm (Van Santen and Buchsbaum, 1997) while ensuring a recording time of less than 150 minutes.
3. Proofread selected sentences to ensure readability of sentences and normalise problematic words (in terms of phonetisation) such as acronyms, abbreviations and foreign language words.
4. Carefully screen and brief voice artists to ensure reading fluency. This was found to be necessary even when working with professional voice artists who sometimes were not comfortable reading text in their first language.
5. Undertake studio recordings with a sound engineer and facilitator keeping track of reading quality.
6. Automatically align phones (Van Niekerk and Barnard, 2009) based on the text-to-speech front-end developed for each language, resulting in sentence, phrase (breath group), word, syllable and phone alignments.
7. Manually verify subsets of alignments where an acoustic distortion measure indicates potential problems in the corpus (Van Niekerk, 2011).

In 2010 we recorded new speech corpora for: Tshivenda, Afrikaans and isiXhosa (see Table 42 for statistics). The Tshivenda corpus was recorded during the first half of 2010 and piloted at Madimbo TSC, Musina, 1-4 June 2010. Feedback from this process was very positive (see section 4.2.2) and the recording protocol used for this language was adopted as part of the effort of improving voice quality.

Afrikaans and isiXhosa were recorded with an amendment to the recording protocol to include English recordings (see isiXhosa secondary recordings in Table 42: Speech corpora

recorded during the project), especially numerical words and dates. This was done to allow further research into building systems that are able to handle simple code-switching (e.g. for the pronunciation of proper names, dates and numbers).

In 2011, TTS speech corpora for English, Sesotho, Setswana, Sepedi and isiZulu were recorded, following the protocol established in 2010. The exception is English where we did not develop source text in-house, instead using text from freely-available English TTS corpora developed at Carnegie Mellon University (2012).

In 2012, we completed the corpus development effort by recording speech corpora for siSwati, isiNdebele and Xitsonga.

Table 42: Speech corpora recorded during the project

Language	Number of utterances	Audio length
Afrikaans	1 005	63 min
English	1 132	61 min
isiNdebele (primary)	994	117 min
isiNdebele (secondary)	395	20 min
isiXhosa (primary)	912	139 min
isiXhosa (secondary)	388	20 min
isiZulu (primary)	918	118 min
isiZulu (secondary)	395	23 min
Sepedi (primary)	1 318	142 min
Sepedi (secondary)	394	23 min
Sesotho (primary)	1 311	129 min
Sesotho (secondary)	395	24 min
Setswana (primary)	1 937	147 min
Setswana (secondary)	395	27 min
siSwati (primary)	973	131 min
siSwati (secondary)	395	23 min
Tshivenda	1 000	83 min
Xitsonga (primary)	942	91 min
Xitsonga (secondary)	395	21 min

11.2.2 Natural language processing

11.2.2.1 Text tokenisation and normalisation

Tokenisation and normalisation routines were initially developed for English and incorporated into Speect in 2010. This functionality was not only extended to Afrikaans in 2011 but the tokenisation routines were also updated to use a regular expression-based algorithm to detect sentence boundaries and word tokens. The regular expression specification of rules makes the segmentation process not only much faster, but also much more robust against difficult cases. The normalisation routine was updated to work hand-in-hand with the tokenisation routine to expand word tokens based on their class. This makes it very flexible to handle ambiguous contexts such as whether “2012” should be expanded as a year to “*twenty twelve*” or as a cardinal number to “*two thousand and twelve*”.

The new tokenisation and normalisation algorithms continue to support custom rule extensions by the user, so that they can be set up for any language. In 2012, we developed these rules for the remaining nine languages.

11.2.2.2 Part-of-speech tagging

Part-of-speech data for the 11 languages were collected and annotated by North-West University (CText) as part of the NCHLT project, and delivered in 2012. Speect was thus extended with a part-of-speech tagger that can train models from these data and tag new sentences at synthesis runtime. Part-of-speech tagging is an important early step in the natural language processing pipeline that is necessary to produce natural synthesised speech from text, as recommended in the research output (Schlünz *et al.*, 2010).

11.2.2.3 Lexical stress and tone

For the Germanic languages, Afrikaans and English, we developed stress-marked pronunciation dictionaries using the rule-based phonetisation schemes for these languages implemented in the *espeak* synthesiser (<http://espeak.sourceforge.net/>). Prototype systems were built with these dictionaries and research on fully exploiting this resource in systems is on-going.

The incorporation of tone information in the Sotho languages required the following:

1. Tone-marked pronunciation dictionaries, providing the *underlying tones* associated with individual words
2. A means of determining *surface tones*, i.e. predicting how *underlying tones* are transformed by linguistic processes such as *sandhi* when words are used in sentence context.

In 2012, we completed the digitisation of an existing Sotho tone-marked dictionary (Du Plessis *et al.*, 1974), providing *underlying tone* information for this language. During the

course of the project, research on a tone labelling algorithm (converting *underlying* to *surface tone*) resulted in the completion of a Master's dissertation (Raborife, 2011), and three published conference papers (Raborife *et al.*, 2010; 2011; 2012). Further integration of these resources into the Sotho TTS system will be the subject of future research and development work.

11.2.2.4 Morphological analysis

North-West University (CText) is still developing morphological analysers for the 11 languages that should be delivered in the first half of 2013. We plan to integrate these analysers into Speect to support other modules where morphological information is needed, e.g. the tone labelling algorithm for Sesotho.

11.2.2.5 Higher-level syntactic and semantic analysis

This functionality is not part of the Lwazi II deliverables, but on-going Master's and Doctoral research is looking ahead at the influences of syntactic and semantic information on the naturalness of speech, in particular emphasis and emotion, using English as a prototype.

11.3 Results

The research and development efforts described in the previous section were ultimately designed to improve the perceived naturalness of TTS systems available in the official South African languages, thereby allowing the use of these systems in real-world applications towards realising the information dissemination goals of the Lwazi II project as a whole.

In this section we measure the success of this effort directly against the proposed goal: to develop systems that are demonstrably more natural than systems developed during Lwazi I. This is done by pairwise comparison of synthesised speech by the old and new systems developed during Lwazi I and II respectively. This process is described below with results presented in Table 43.

For these tests respondents were asked to listen to pairs of utterances from a test set. Pairs of utterances, A and B, were synthesised from a single sentence (text), using the Lwazi I and II voices respectively. Respondents were then prompted to choose which utterance sounds more natural or whether both have a similar quality. Each respondent did this for 20 to 30 test sentences that were fabricated simple statements in the particular language.

The voice accredited with the most test sentences was then deemed to be more natural. McNemar's test was then applied to determine the validity of the results. McNemar's test is a chi-square test for paired sample data that is used to examine the statistical significance of results. Its chi-square value is greater than or equal to 3.841 for a significant outcome. If the value is less than 3.841 the outcome is not significant. For a fuller exposition on McNemar's test, see (Schlünz *et al.*, 2010).

The results are listed in Table 43. The first column lists the language and the second column the total number of utterances evaluated. Columns 3 and 4 list the number of utterances accredited to each voice and column 5 the number found equal. The last column lists the McNemar chi-square scores for significance (which are independent of the equal counts). From these results it can be seen that, for all 11 languages, the Lwazi II voices clearly outperform their Lwazi I counterparts significantly in terms of naturalness.

Table 43: Perceptual test results

Language	Total	Lwazi I	Lwazi II	Equal	Chi-square
Afrikaans	200	24	174	2	112.880
English	200	10	179	11	150.224
isiNdebele	353	51	286	16	163.176
isiXhosa	190	53	134	3	34.654
isiZulu	506	62	398	46	244.696
Sepedi	655	197	335	123	35.538
Sesotho	200	56	133	11	30.964
Setswana	200	30	165	5	92.771
siSwati	370	75	236	59	82.830
Tshivenda	200	31	160	9	86.452
Xitsonga	444	87	352	5	159.363

11.4 Summary of outputs

In this section we summarise the tangible outputs directly associated with this project. This is divided into **software** and **data**. The TTS publications have been captured in section 16.3.

11.5 Software

- The Speect speech synthesis platform (available at <http://speect.sourceforge.net>), featuring:
 - Mixed excitation HMM-based synthesis by extending the HTS engine (available at <http://sourceforge.net/projects/me-hts-engine>)
 - Portability to Linux, Windows and Android operating systems
 - Extensive documentation including user and programming guides
 - Specialised NLP modules for South African languages
 - Experimental support for polyglot (multilingual) speech synthesis
 - Integrated scripts for phone alignment and acoustic modelling using HTS.
- Prototyping tools for TTS development (available at <https://github.com/demitasse/ttslab> and <https://github.com/demitasse/ttslabdev>).

11.6 Data²¹

- Speech corpora with phone alignments and technical documentation.
- Stress/tone-marked pronunciation dictionaries for English, Afrikaans and Sesotho with technical documentation.
- Modules for Speect representing the TTS voices for the 11 official languages as evaluated in this report.

12. Automatic speech recognition (ASR)

12.1 Context and goals

The automatic speech recognition (ASR) systems developed as part of Lwazi I (2006-2009) are more than 90% accurate when used for “small vocabulary” tasks, that is, when determining which of a set of 10-20 keywords have been spoken. Such systems are usable for very specific tasks within a spoken dialogue system (SDS), for example when a user is choosing among a limited set of options. However, ASR systems really start adding value to SDSs when they can handle larger vocabularies, for example, when a user can access information by naming a specific town or school.

The main goal of the ASR component of the Lwazi II project (2010-2012) was therefore to improve the accuracy of the ASR systems in all 11 official languages, so that these systems are able to achieve average accuracies of greater than 80% for vocabularies of 100-200 words.

In addition, the following sub-goals were important:

- Developing students skilled in research and development of multilingual ASR systems.
- Contributing new knowledge on ASR relating to the South African languages to the international HLT community.
- Developing improved and extended linguistic resources, including larger speech corpora and more accurate acoustic models, phone sets and pronunciation dictionaries, as required, to reach the main goal.

In 2010, the main focus of the ASR team was on obtaining experience in system optimisation and on developing the necessary optimisation tools required. Four languages were chosen to conduct initial optimisation experiments. In 2011, optimised phone recognition systems were implemented for the seven languages that were not included in the experiments conducted during 2010. In addition, initial experiments were conducted on tasks of increasing difficulty for four selected languages. During 2012, ASR systems that are able to

²¹ All data was transferred to DAC on 4 January 2012.

achieve the goal of average accuracies of greater than 80% for vocabularies of 100-200 words were developed in all 11 languages. Longer term research contributing towards a better understanding of ASR for South African languages continued and the student development programme also continued throughout the course of the project. The goals for the project, as summarised in the project overview above (section 1.3.2), were all successfully met by the end of the project.

The progress made during the course of the project is described in the following sections:

- Section 12.2 addresses the **main longer term research themes**. This research contributed to an improved understanding of ASR for the South African languages, but did not always lead directly to improved systems. The techniques will be incorporated into next-generation systems as they mature.
- Section 12.3 discusses student training and related **HCD** activities.
- The method that was used to implement **optimised phone recognition accuracy** in all eleven languages and the corresponding results are presented in section 12.4.
- A report describing the **improved density estimation** that was developed to achieve the goal of average accuracies above 80% for vocabularies of 100-200 words is attached as annexure H at the end of the report.

12.2 Main research themes

The main ASR research themes all address challenges encountered when developing ASR systems for the South African languages. Immediate outputs include the publications listed in section 16.3.2, and referenced below.

12.2.1 Modelling code-switched speech

Speech recognition corpora often contain code-switched speech: utterances that contain words in more than one language. For example, even when constrained to a spoken dialogue, many speakers of South African languages would use English numbers, dates and times. In addition, many place names have pronunciations that are clearly linked to other languages spoken in the vicinity. Very little is known about the extent to which code-switching happens in South African languages, or what the best approaches are for identifying and modelling these words.

Over the course of the project we aimed to develop the necessary tools and techniques to deal with code-switched speech when developing ASR systems, with our initial work focussing on the Sotho languages. An initial exploratory study of this topic using the Lwazi I corpus was completed in 2010 (Modipa and Davel, 2010). More recent results were published in 2012 (Modipa, *et al.*, 2012). A prompt set was subsequently designed to elicit code-switched speech from Sepedi speakers. The design was based on transcripts of radio broadcasts that included various examples of code-switching. Spoken versions of the

prompts were collected using Woefzela (CSIR Meraka Institute, 2011) and are currently being used to compile a database containing examples of code-switching in Sepedi.

12.2.2 Improved pronunciation modelling

Until the pronunciation dictionary development work in the NCHLT Speech project was initiated, the Lwazi I pronunciation dictionaries were the most comprehensive electronic pronunciation resource available for South African languages. These dictionaries are fairly small (at 5 000 words) and only contain generic language-specific words. They form a subset of the much larger vocabularies available in these languages, and also exclude all proper names.

Accurate pronunciations of South African proper names are not readily available. (Such large onomastic corpora do exist for many of the major world languages.) The multilingual nature of South African society makes this task even more formidable, with different language communities pronouncing names originating from other language communities in different ways. We therefore aimed to develop more accurate pronunciation models of the South African languages, specifically by –

- investigating cross-lingual pronunciations of South African proper names (Kgampe and Davel, 2010), and whether there are systematic effects that we can model in a data-driven fashion;
- continuing to develop larger pronunciation dictionaries, and improving the tools used during pronunciation dictionary development (Davel and De Wet, 2010); and
- experimenting with different phoneme sets, and continuing to optimise these for ASR purposes (Modipa *et al.*, 2010).

Results from the study on cross-lingual personal name pronunciation indicated that South African personal names show large variability in pronunciation, especially when users are unfamiliar with the names themselves: only one in three words are pronounced “correctly” when produced in a cross-lingual fashion. Even familiar names show large variability, with about one in two pronounced correctly. This affirms the importance of including variants for proper names in general. The results also showed that many of these variants are fairly predictable, which should facilitate the automatic generation of these variants under very specific conditions – an ongoing topic of study within the HLT RG (Kgampe and Davel, 2011).

12.2.3 Data combination and sharing

As more data becomes available (also from the projects commissioned by the South African National Centre for HLT), research relating to the combination of data sources is becoming more and more pertinent. In this area the following was investigated:

- Channel normalisation techniques, whereby the mismatch between corpora collected under different recording conditions were analysed and compensated for.

- Cross-lingual data combination, where strategies were investigated for sharing data across languages. For this purpose, tools to analyse the distances between language pairs (and specific phonemes across languages), are important.

Whether data are combined from different sources, or from different languages, the type of combination scheme (including various forms of data adaptation) provides a fertile research environment and only preliminary results have been produced on this topic (Van Heerden *et al.*, 2010).

Even though it would be ideal to train acoustic models using environment-specific audio data, it is possible to use mismatched audio data to train domain-specific acoustic models and create a robust system by using a combination of environment normalization and adaptation techniques. For instance, given access to large volumes of high-bandwidth audio data, it is possible to create telephony-based applications without having to collect telephony recordings. Similarly, given access to audio data sourced from varying environments, after data normalization, the audio data may be pooled to create a relatively large training corpus. To further improve the robustness of an automatic speech recognition system, relatively small amounts of audio data can be used to rapidly adapt the acoustic models to model the channel mismatch and compensate for the mismatch (Kleynhans *et al.*, 2012).

The experiments described in annexure H were all conducted using high-bandwidth audio data but, given the possibilities of the environment normalization and adaptation techniques mentioned in the previous paragraph, the results are equally applicable to system development in a telephony environment.

12.2.4 Trajectory modelling

While current speech recognition techniques are highly successful when presented with very large training corpora, the reason why such large corpora are necessary is not clearly understood. In this work, we analyse the acoustic variation introduced by co-articulation from first principles, in order to understand this effect better (Badenhorst *et al.*, 2010). The analysis may also enable us to more accurately model the effect of co-articulation for speech (Badenhorst *et al.*, 2011). Our end goal is the development of techniques that make more efficient use of limited data. While this is one of our more ambitious research themes, initial results are promising (Badenhorst *et al.*, 2012).

12.3 ASR HCD

A number of students participated in the Lwazi II project by conducting post-graduate research in ASR. Details with regard to these students – Charl van Heerden, Jaco Badenhorst, Mpho Kgampe, Neil Kleynhans and Thipe Modipa – are summarised in section 16.2.

Additional HCD activities included:

- A research visit hosted for 13 post-graduate students from the University of Limpopo in September 2010. Coordinated by Thihe Modipa at Meraka, and facilitated by Jonas Manamela at University of Limpopo, the research visit was aimed at stimulating student interest in post-graduate speech technology research. The students, all busy with an honours programme in Computer Science at the time, were provided with an overview of the key aspects of the Lwazi II project, including speech technology applications, speech technology system development, TTS, ASR and ASR data collection. Technical presentations and a showcase of deployed speech technology applications completed the visit.
- A site visit hosted for a group of post-graduate students from the North-West University took place in September 2010. The students spent two days at Meraka. On the first day Jaco Badenhorst, Neil Kleynhans and Thihe Modipa presented a tutorial on the ASR tools developed at HLT during 2010 and provided the students with a scripting backbone that enabled them to build their own ASR systems. The students worked in groups and each group worked on a different language. After they returned to campus, the students continued to participate in one of the Lwazi II sprints under the supervision of Charl van Heerden and Febe de Wet. They actively contributed to the results presented in Table 44.
- In April 2012 the HLT group at Meraka hosted a five-day HLT Autumn School. The school was attended by 26 students from Stellenbosch University (2), University of Limpopo (10), North-West University (8), DAC (1) and the HLT group at Meraka (5). Lectures were presented by a number of speech scientists from different universities and organizations in South Africa and covered topics like pattern recognition, an introduction to linguistics, ASR, TTS, etc.

12.4 ASR system optimisation

The ASR research that was conducted during Lwazi II was done in two phases. Phase 1 involved the optimisation of **phoneme accuracy** for systems trained on Lwazi I data. During phase 2, data that was collected during the NCHLT Speech project (CSIR Meraka Institute, 2012) was used to train systems capable of **word recognition** on vocabularies of 100-200 words with more than 80% accuracy. Both phases of the work involved all 11 languages. In this section we describe the process that was used to optimise **phoneme accuracy** using the Lwazi data. The development of the systems that are capable of performing recognition tasks of increased complexity are described in annexure H.

12.4.1 Optimised phoneme recognition

The following paragraphs describe the phoneme recognition optimisation process. Firstly, a motivation is given for the **evaluation protocol** that was chosen to gauge system performance. This is followed by a description of the changes that were made to the **ASR**

system development toolkit. A number of **optimisation techniques** were applied in order to improve the accuracy of the acoustic models, which were implemented to yield the **results**.

12.4.1.1 Evaluation protocol

The evaluation protocol defined a fixed evaluation subset for the Lwazi corpus. The evaluation set – all utterances from 20 male and 20 female speakers – was formally removed from the training set and was only used for reporting results. This means that less data was available for acoustic modelling, but that results could be tracked very systematically over the course of the project.

Results are reported in terms of unconstrained phoneme recognition: here, the system assumes any sequence of sounds could have been spoken (not just allowed combinations of sounds). This is an artificial task, but it provides a rigorous way of measuring the improvement of acoustic models.

When phoneme recognition performance is evaluated, the number of phonemes in a recognised string may be different from the number of phonemes in the reference string (what should have been recognised). For this reason, two measures of performance are used: correctness and accuracy, where:

$$\text{Correctness} = \frac{\text{Correct labels}}{\text{Total labels}} \times 100\%$$

and

$$\text{Accuracy} = \frac{\text{Correct labels} - \text{Inserted labels}}{\text{Total labels}} \times 100\%$$

12.4.1.2 ASR system development toolkit

The ASR system development toolkit was modified significantly during the course of the project to make system experimentation and optimisation more efficient. It was also extended so that key tasks (such as decoding) could run in parallel. During 2010, the new toolkit was extensively tested through a month of “sprints”: each sprint was a two-week period during which an entire team focused on a single activity, meeting every day to gauge progress. The team consisted of six students and three full-time researchers. By the end of the second sprint, the toolkit had been tested thoroughly and a significant amount of knowledge transfer had taken place, with the entire team trained in the use of the toolkit.

12.4.1.3 Optimisation techniques

Apart from the improved toolkit for ASR system development, the following techniques proved to be most effective in improving the phoneme accuracy of the ASR systems derived from the Lwazi I data:

- **Improved speaker normalisation:** Various techniques can be applied to normalise speech prior to acoustic training. This means that an individual’s speech is processed to first seem more “neutral” (speaker-specific effects are removed as far as possible) before combining all data and training models. The largest win was obtained using speaker-based cepstral mean and variance normalisation (CMVN), which was subsequently applied to all systems.
- **Improved pronunciation modelling:** During the course of the project, a number of specialised pronunciation dictionaries were developed, some also as part of other projects (Davel and De Wet, 2010). Selected pronunciations from these dictionaries were used to extend the Lwazi I dictionaries, improving performance. In addition:
 - The idea to model affricates in the Sotho languages as a combination of the constituting sounds, as described in (Modipa *et al.*, 2010), was implemented to improve pronunciation modelling for **all three Sotho languages**.
 - **English** pronunciation modelling was improved by extending the Lwazi I South African English (SAE) pronunciation dictionary with a so-called “background” dictionary. This dictionary contains hand-crafted pronunciations for words that are conflicting with the grapheme-to-phoneme rules that are generally used to create pronunciation dictionaries for the Lwazi systems. In addition to ordinary words, the background dictionary contains examples of proper names and acronyms, (e.g. Johannesburg and Fifa), abbreviations and initialisms (e.g. sms, CNN and CSIR), compounds with initialisms (e.g. M-Web and Cell-C), informal words (e.g. merc – Mercedes and bru – brother) and number words (e.g. 4x4, media24 and 94.2).
 - Pronunciation modelling in **Afrikaans** was improved by means of a similar background dictionary and by using a much more extensive pronunciation dictionary that was developed in another project (Davel and De Wet, 2010).
 - The **isiZulu** pronunciation dictionary was improved by manually verifying the pronunciations of the 200 words that occur most frequently in the transcriptions. A similar approach was attempted for the other languages as well, but the number of erroneous pronunciations found either did not have an impact on recognition accuracy or the pronunciations could not be verified by a first language speaker of the language.

12.4.1.4 Results

The optimisation results presented in this section are quantified in terms of the improvement in unconstrained phoneme recognition accuracy. While these results cannot be interpreted as intuitively as word recognition accuracy, they provide a rigorous measurement of performance improvement.

Three different sets of results are presented in this section. The first set is presented in Table 44. Results obtained for the baseline systems are compared with the systems improved by implementing CMVN. Both phoneme correctness and phoneme accuracy are reported. The results in Table 44 show that, on average, CMVN leads to a reduction in error rate of more than 3%.

During the transcription of the Lwazi II data, a convention was established to indicate non-speech events such as speaker and background noise. These labels were used to determine the recognition accuracy for the systems with CMVN using only the utterances in the test set without any noise labels, i.e. the clean data only. These results are presented in Table 45. The percentage difference in error rate is given as an indication of the negative impact that the noisy data has on system performance. Table 45 shows that, on average, the difference in error rate between the complete test sets and the clean subsets of the test sets is more than 3% for phoneme correctness and more than 8% for phoneme accuracy. This observation clearly illustrates the negative impact that noisy data has on system performance.

Table 44: ASR results for all 11 languages

Language	Baseline		CMVM		% reduction in error rate	
	Correctness	Accuracy	Correctness	Accuracy	Correctness	Accuracy
Afrikaans	68.40	58.19	69.30	58.95	2.85	1.82
isiZulu	67.39	55.34	68.59	57.24	3.68	4.25
SAE (English)	60.17	49.19	61.17	50.59	2.51	2.76
Sepedi	73.45	64.14	74.22	65.02	2.90	2.45
Sesotho	73.69	63.86	74.65	64.91	3.65	2.91
Setswana	73.06	62.79	74.09	63.83	3.82	2.79
isiNdebele	73.91	65.23	74.98	66.23	4.10	2.88
Xitsonga	72.01	62.30	73.37	64.54	4.86	5.94
isiXhosa	74.32	64.48	75.13	65.65	3.15	3.29
Tshivenda	75.29	66.08	76.67	67.93	5.58	5.45
siSwati	74.70	65.55	74.87	66.01	0.67	1.34
Average	71.49	61.56	72.46	62.81	3.43	3.26

Table 45: ASR results for all 11 languages on clean subset of test data only

Language	CMVN – all data		CMVN – clean data only		% difference in error rate	
	Correctness	Accuracy	Correctness	Accuracy	Correctness	Accuracy
Afrikaans	69.30	58.95	71.06	61.97	5.73	7.36
isiZulu	68.59	57.24	69.71	61.39	3.57	9.71
SAE (English)	61.17	50.59	62.68	53.58	3.89	6.05
Sepedi	74.22	65.02	74.86	67.79	2.48	7.92
Sesotho	73.69	63.86	74.22	66.94	2.01	8.52
Setswana	74.09	63.83	76.42	68.71	8.99	13.49
isiNdebele	74.98	66.23	75.61	68.94	2.52	8.02
Xitsonga	73.37	64.54	74.27	66.67	3.38	6.01
isiXhosa	75.13	65.65	76.10	68.49	3.90	8.27
Tshivenda	76.67	67.93	76.97	69.76	1.29	5.71
siSwati	74.87	66.01	74.63	68.55	-0.96	7.47
Average	72.37	62.71	73.32	65.71	3.35	8.05

The third set of results, shown in Table 46 below, illustrates the additional gain in recognition accuracy that was achieved by improving pronunciation modelling in SAE, Afrikaans, isiZulu and the Sotho languages²².

Table 46: ASR results for systems with improved pronunciation modelling

Language	CMVN		CMVN + pronunciation modelling		% reduction in error rate	
	Correctness	Accuracy	Correctness	Accuracy	Correctness	Accuracy
Afrikaans	69.3	58.95	69.63	59.71	1.07	1.85
isiZulu	68.59	57.24	69.07	57.41	1.53	0.4
SAE (English)	61.17	50.59	60.43	49.64	-1.91	-1.92
Sepedi	73.19	61.62	75.31	65.82	7.91	10.94

²² The results for the Sotho languages in Table 44 and Table 46 are not exactly the same because the number of phonemes in the two systems is not the same: the systems which model the affricates as a combination of individual phones have fewer phone models than those that have affricate models. If results are compared in terms of phoneme recognition accuracy, an adjustment should be made for the difference in the number of phones. The results shown in Table 46 therefore correspond to the adjusted CMVN results presented in Table 44.

Language	CMVN		CMVN + pronunciation modelling		% reduction in error rate	
Sesotho	72.74	61.54	74.81	65.59	4.07	10.53
Setswana	72.78	61.02	75.01	64.50	8.19	8.93

The results in Table 46 show that the alternative pronunciation dictionaries improved system performance for all the languages that were evaluated, with the exception of SAE. The deterioration in performance for the English system can probably be explained by the fact that the words modelled in the background dictionary do not occur frequently in the Lwazi data. The biggest gain in system performance is observed for the Sotho languages for which affricate splitting was implemented in the alternative pronunciation dictionaries.

12.5 Conclusion

The outcomes as intended during the ASR research and development component of Lwazi II were attained. The main achievements are summarised in Table 47.

Table 47: ASR research and development goals

Goals	Deliverables 2010 – 2012
Improved accuracy	> 80% on 100-200 words, all languages
HCD	Post-graduate students trained
New knowledge	Publications with regard to ASR for SA languages
Linguistic resources	Additional resources developed during project packaged and released

13. Platform development

13.1 Speech processing pipeline

13.1.1 Description

This section reports on the speech processing pipeline developed during Lwazi II. The speech processing pipeline was developed to support the speech recognition goals of the Lwazi II project, including:

- The collection of larger data sets for certain languages
- Enhanced recognition accuracy by cross-language sharing of data and/or models
- Improved acoustic models, phone sets and pronunciation dictionaries through statistical modelling and error analysis of our current systems.

13.1.2 Requirements

The following requirements were developed for the speech processing pipeline:

- The ability to buffer incoming speech audio
- The ability to manage speech recognition resources (e.g. models)
- The ability to handle multiple recognition requests simultaneously
- The ability to administer the speech decoder (e.g. Julius)
- The ability to return recognition information to the calling application, including a confidence metric.

13.1.3 Design

A speech server is the speech processing pipeline's entry point for speech audio. Audio is sent to the speech server by the calling application (e.g. Asterisk) and speech results are sent back via a client/server communication protocol. Upon receiving speech audio, the speech server administers the following steps:

Step 1: End Pointer

The end pointer determines when speech starts (i.e. when the signal being received has enough energy in the frequency bands one would usually associate with speech) and also when speech has ended. This signals the decoder when to start decoding a sentence and also when to conclude the decoding process.

Step 2: Mel Frequency Cepstral Coefficients

Mel frequency cepstral coefficients are the features that are extracted from the raw speech signal. Each mel frequency cepstral coefficient is represented as a 13 dimensional vector, which is calculated over a 25 ms window of speech. MFCCs are calculated every 10 ms.

Step 3: Cepstral Mean Subtraction

Cepstral mean subtraction is performed to remove channel and gender biases from the signal. A default mean value is used for the first couple of frames, where after this mean is updated to better reflect the true mean of the signal.

The default mean needs to be estimated from the speech though. In an offline environment, this is simple, as the mean is calculated on the entire file (static CMN) and then subtracted. In a real-time environment where speed is a critical factor, the process is typically initiated with some pre-calculated mean which is then systematically adapted to the incoming speech as more and more frames are seen by the decoder (real-time CMN).

Julius systematically “adapts” the initial mean vector by subtracting a weighted version of the initial mean from each feature with the following formula

$$c_j = c_j - \left(\left(\sum_{i=0}^j c_i \right) + u_{init} \cdot \omega \right) \cdot (j + 1.0 + \omega)^{-1}$$

where $w=100$ and u_{init} is a user-specified initial mean.

Step 4: Deltas

First order differential coefficients are calculated by subtracting adjacent MFCCs from one another.

Step 5: Double Deltas

Second order differential coefficients are calculated by subtracting adjacent first order differential coefficients from one another. One can save one “time frame” by calculating first and second order differentials at the same time as opposed to a serial setup.

Step 6: Acoustic Models

The acoustic models are trained using HTK and are 3-state, 39 dimensional Gaussian Mixture models. Distributions are shared where it makes sense (tied mixture system).

Step 7: Language Models

These can either take the form of a small to medium vocabulary task where terms may have equal probability, to n-gram models which are trained on large text corpora.

Step 8: Decoder

The decoder takes as input frames, an acoustic model as well as a language model and provides as output a decoded phoneme (or word) sequence, together with associated acoustic scores, which reflects how well the frames fit the best possible output token sequence.

Step 9: Text Output

The system creates a decoded textual representation of the speech signal that was received.

After the textual representation of the speech signal has been generated, the speech server passes the data back to the calling system which will use the information to perform appropriate application logic.

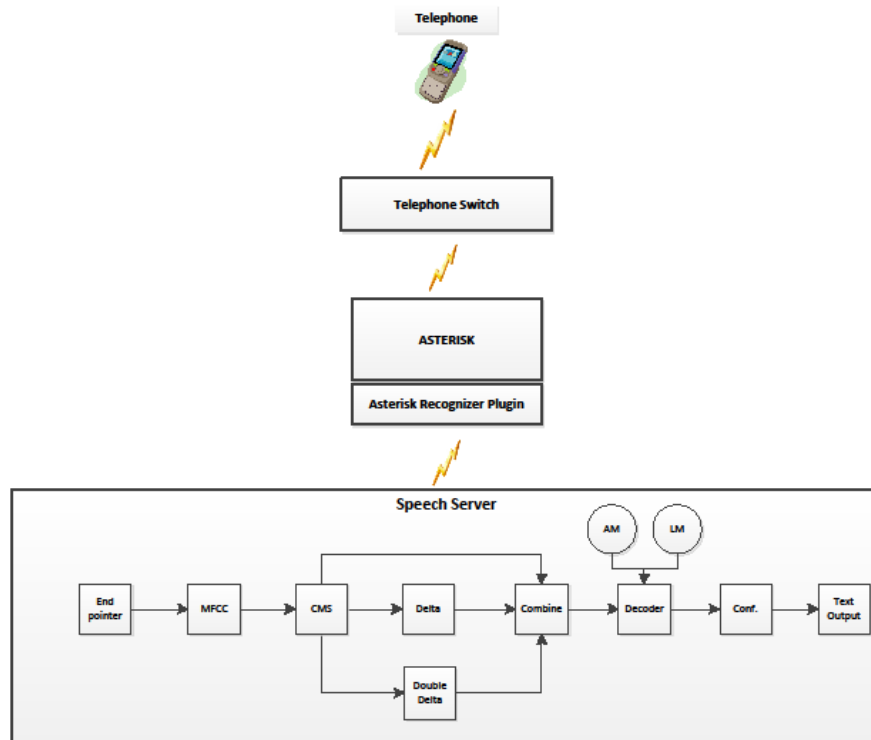


Figure 40: Speech processing pipeline design

13.1.4 Implementation report

The second development cycle resulted in the completion of the following required features:

- Removing dependency on proprietary software licenses:
 - Adoption of open-source Julius decoder
 - Integration of speech models into Julius decoder
 - Tuning Julius for optimal recognition results.
- Implementing a simple test system for user evaluation.
- Developing supporting documentation.

Each of these tasks was completed successfully resulting in an end-to-end speech pipeline ready for use by an application.

13.1.5 Testing

The pipeline was tested to verify three important aspects:

1. Performance should be comparable to HTK's Hvite.
2. Performance in real-time mode should be comparable to offline mode.
3. Real-time performance should be improved when the features are extracted using a real-time mean instead of a static mean.

As a recognition task, a 200-term grammar was defined by randomly selecting 200 terms from an in-house corpus. These terms collectively occur 5 723 times and cover 200 different speakers. In order to recognize all 5 723 terms without any speaker overlap between training and evaluation sets, 4-fold cross-validation was employed.

The evaluation task was thus to decode incoming speech, with the output being one of 200 possible terms. This is a reasonable simulation of a medium vocabulary recognition task.

A baseline was established by performing static CMN on broadband data (16kHz sample rate) and performing ASR using HVite. This was compared to decoding the same feature set using Julius.

Real-time broadband data was then recognized using the same static CMN models as mentioned above.

Lowband data (8kHz sample rate) was then normalized using static CMN and used to train a model. Recognition was again performed; both with static CMN normalized data, as well as real-time normalized data. A model was then trained using real-time normalized data and used to recognize real-time normalized data.

Detailed results are presented in annexure J.

Table 48: 200-term recognition accuracy for various decoders and feature sets

Model	Data	Decoder	Feature type	Accuracy
Static CMN, 16kHz	Static CMN, 16kHz	Hvite	MFCC	98.76
Static CMN, 16kHz	Static CMN, 16kHz	Julius	MFCC	98.31
Static CMN, 16kHz	Real-time CMN, 16kHz	Julius	MFCC	97.89
Static CMN, 8kHz	Static CMN, 8kHz	Julius	MFCC	98.08
Static CMN, 8kHz	Real-time CMN, 8kHz	Julius	MFCC	97.78
Real-time CMN, 8kHz	Real-time CMN, 8kHz	Julius	MFCC	98.03
Real-time CMN, 8kHz	Real-time CMN, 8kHz	Julius	RAW	98.03

13.1.6 Discussion

The baseline results are comparable and serve as a verification that Julius performs as well as HTK on the decoding task. A drop in performance is observed when decoding real-time CMN data with acoustic models trained on static CMN data. The same trend is observed when low-band data is used, with an expected overall drop in accuracy when compared to high-band data.

When recognizing real-time CMN data with real-time CMN models, the system performs much better than when trying to recognize real-time CMN data with static CMN models. However, it does not perform quite as well when recognizing static CMN data with static CMN models.

As a final sanity check, raw audio data was decoded with Julius being forced to run in real-time mode, with the result virtually identical to the case when feature extraction was performed beforehand.

13.1.7 Looking forward

Future revisions of the ASR pipeline might focus on the following features:

- Performance tuning
- Improved ASR resource sharing
- Automating project build.

13.2 Content management system

13.2.1 Description

This section provides a report on the content management system created for Lwazi II. The Lwazi content management system was developed to support quick and easy creation of speech-enabled applications for the Lwazi II project.

13.2.2 Requirements

The requirements for the content management system were as follows:

- The ability to manage content using a language-independent, distributed, web-based management and maintenance system.
- The ability to administer and configure IVR and ASR systems with a single unified tool and language-independent configuration files.

13.2.3 Design

A web management interface is used to easily update the application database. The data model drives the application's published website and IVR content. This web management layer requires only application-specific forms to create new applications. Interfaces in additional languages can be added by inserting translated texts.

On the IVR server, a control interface manages the IVR and ASR systems using a single authoritative command interface. Language-independent configuration is possible using application scripts.

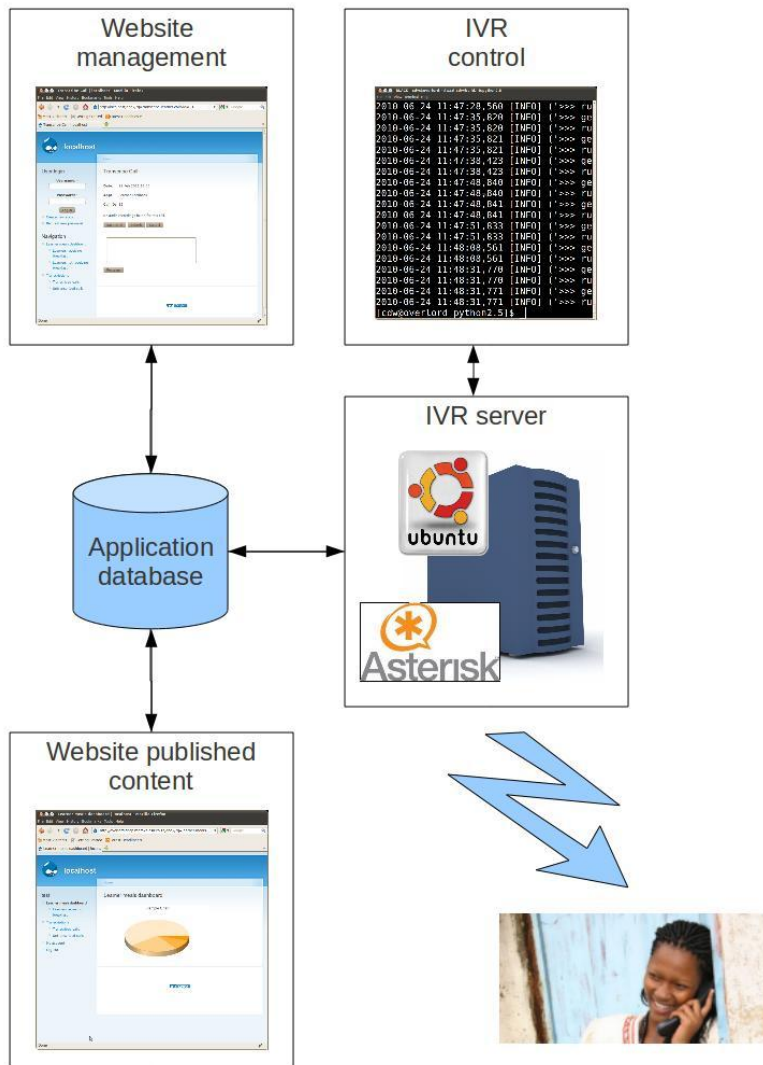


Figure 41: Content management system design

13.2.4 Testing

The Lwazi content management system was tested by developers, the application team, and finally by pilot testers at the application deployment sites. Additionally, improvements were made through feedback from existing system deployments.

13.2.5 Progress

Version 2.0 of the content management system provided the following improvements:

1. Greater IVR control from the website
2. Improved IVR statistics on the website
3. Improved IVR content authoring on the website.

13.3 Telephony platform improvements

13.3.1 Platform run time testing

The new speech processing pipeline (section 13.1) was not operationalised in the deployed services, and therefore its run time could not be tested. Instead, we analysed the up and down time of the operationalised services, to be able to identify and address problems.

Service up and down time information for the DBE Learner IVR application is shown in the following figure. As depicted, the service was up for a total of 493 days from August 2011 through to October 2012. The longest consecutive period of uptime without any failures was 70 days from January to February 2012 – amounting to 1 680 hours (70 days x 24 hours). The majority of service downtime can be attributed to scheduled application upgrades and maintenance, as well as power failures at our internal hosting site within the CSIR. Uptime for newer services, such as Afrivet, has been improved significantly due to the use of external hosting solutions that can guarantee 24-hour service uptime with dedicated support staff.

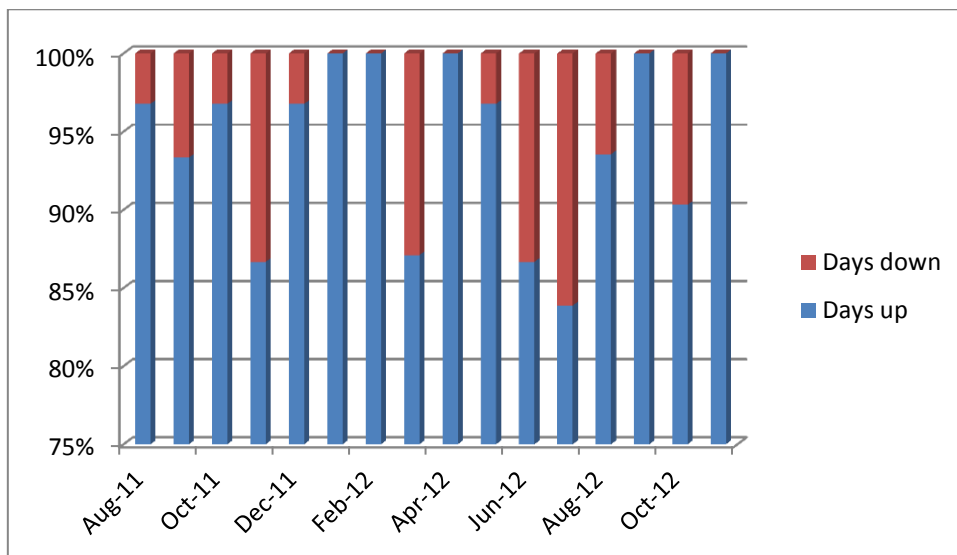


Figure 42: DBE Learner IVR application up and down time

13.3.2 Externally hosted infrastructure

13.3.2.1 Description

All of the telephone and web-based services developed for Lwazi II (the CDW, DBE, and Afrivet services) require replicable telephony and computing infrastructure. These services rely on the availability of telephone lines, SMS gateways, as well as computer servers for running telephony platform and application software, as well as web frameworks and applications.

13.3.2.2 Motivation

Most of the services developed for this project were deployed by hosting hardware and software internally using the HLT research group's own resources. This, however, introduced a number of challenges over time, including:

- Running out of outgoing lines in our building (expensive hardware would need to be purchased to increase this capacity)
- High cost of leasing sufficient outgoing lines
- Cost of purchasing and maintaining internal hardware (computer servers, telephony hardware)
- The inability to guarantee 24-hour uptime (no support staff available).

The majority of these challenges could be addressed by approaching external hosted service providers for any new services developed.

13.3.2.3 Solution design

Hosting of interactive voice response and web services requires both telephony and computing hardware. On the telephony end there is a need for telephone lines to enable incoming and outgoing calls, as well as SMS gateways for sending text messages. Computer servers are also required to run the Lwazi platform and application software.

Three external service providers were identified to meet this need; the first one capable of providing the required telephone lines, the second one capable of providing outgoing SMS service hosting, and the last one capable of providing servers to host our software services. Switchtel (www.switchtel.co.za/) was selected as our telephony provider, BulkSMS (www.bulksms.com) as our SMS provider and Snowball (www.snowball.co.za/) as our hosted server provider. The following figure is a high level view of how the service is hosted externally.

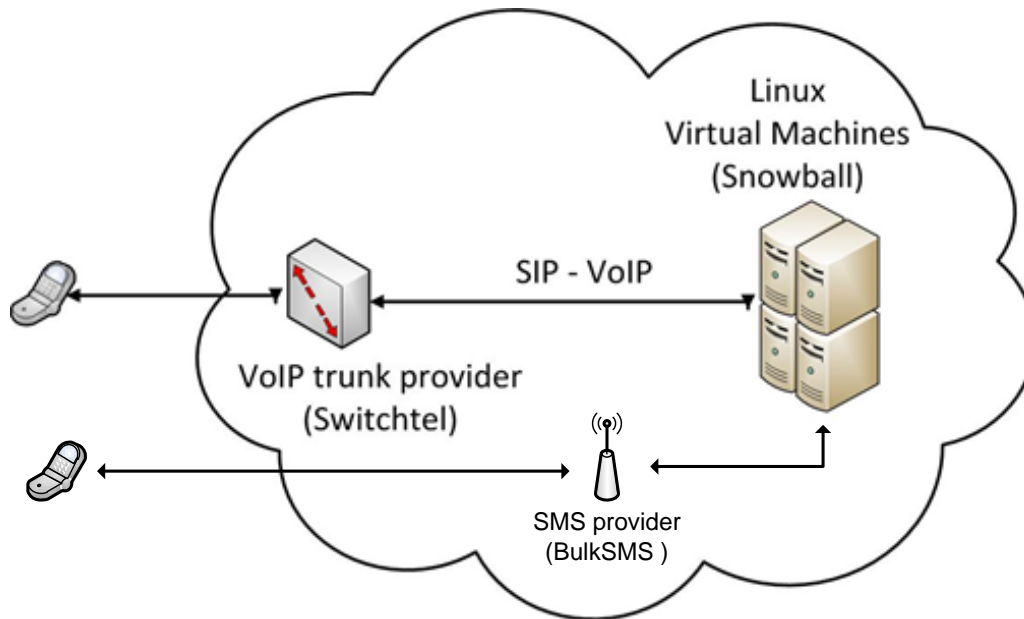


Figure 43: External hosting configuration

As depicted, phone calls get made via a voice over internet protocol (VoIP) trunk provided by Switchtel. These calls then get forwarded to our software services running on hosted server virtual machines provided by Snowball. In turn, SMS messages get sent out to users via the SMS provider, namely BulkSMS.

13.3.2.4 Testing

The entire Afrivet application (all services) was deployed on the above external hosting infrastructure. IVR services were rolled out using two service providers, firstly Switchtel who provided physical telephone lines and VoIP trunking, and secondly Snowball Effect who provided computer servers to host the telephony platform and application software. The web service was also provided using hosted computer servers provided by Snowball. Lastly, outgoing SMS was enabled by the use of an external SMS service provider, namely BulkSMS. Use of these external service providers enabled –

- better service scalability to handle larger scale usage,
- more efficient costing since no expensive computing and telephony hardware had to be purchased, and
- better service uptime since external providers could guarantee 24-hour service uptime.

14. Technology evaluation

The terms of reference for the project set out the following technology evaluation criteria. In most cases, the target measurements were either achieved or exceeded. There were two exceptions:

1. The target for the number of interns to be employed was four but only one intern was appointed. Two staff members, who were interns at the start of the project, were appointed as permanent employees. In addition, three CSIR Meraka staff members from other research groups joined the project for various periods of time and were exposed to the research involved. The target for post-graduate students enrolled with topics related to the project, was exceeded by nine. Three students, who graduated with topics related to the project, were appointed as permanent CSIR employees.
2. The targets for the number of local and international journal publications were three and two respectively, while only one local and one international journal article were published. This was, however, supplemented with one book chapter and 18 more conference papers than required.

Table 49: Technology evaluation criteria

Major criteria		Evaluation	
Criterion	Target measurement	Achieved (Y/N)	Reference
Success of deployment	Deployed services have been running for at least two months at selected sites.	Y	Sections 7 & 8
Quality of open technology components developed	a) ASR systems: at least 80% word recognition accuracy for 100-200-word vocabularies. b) TTS systems: voices that are demonstrably more natural-sounding than the Lwazi voices (measured through pair-wise testing).	a) Y b) Y	a) Section 12.4.1.4 b) Section 11.3
Stability, speed and robustness of speech processing pipeline	a) Mean time before failure of at least 200 hours under production conditions. b) Speed of the speech recognition component sufficient to perform recognition at real time for realistic grammars using vocabularies of up to 200 words.	a) Y b) Y	a) Section 13.3.1 b) Sections 13.1.5 & 13.1.6
Diversity, number and quality of trained	HLT skills developed in speakers of official languages; at least 4 interns	N	Sections 1.4; 9; 12.3; 16.2

Major criteria		Evaluation	
Criterion	Target measurement	Achieved (Y/N)	Reference
human resources	trained during the project; at least 4 post-graduate students enrolled with research topics that support the project.		
Availability of open speech resources produced	All resources supporting ASR and TTS systems packaged, documented and made available to the research community.	Y	Section 15
Quality of scientific work produced	At least 3 publications submitted to local journals, 2 publications submitted to international journals, and 5 publications submitted to local/international refereed conferences.	N	Section 16.3
Time to market	Final deployed services start production between 24-30 months of project initiation.	Y	Sections 7.4 & 8.4

15. Technology release²³

15.1 High-level technology overview

Figure 44 provides a high-level overview of the technologies and applications developed by the HLT RG over the project period.

Resources, tools and technologies developed are indicated in dark purple, with the sub-technologies indicated in a lighter purple. Woefzela and DictionaryMaker are indicated in green as these technologies were not developed within the Lwazi II project. Woefzela is included in the high-level overview as it forms an integral part of resource (speech) collection. Likewise, DictionaryMaker, developed within the Lwazi I project, is shown as it is needed for pronunciation dictionary development and thus forms part of the speech technology development process.

²³ A comprehensive technology transfer master plan will be delivered to DAC as a stand-alone deliverable. The information presented here has been extracted from the Plan.

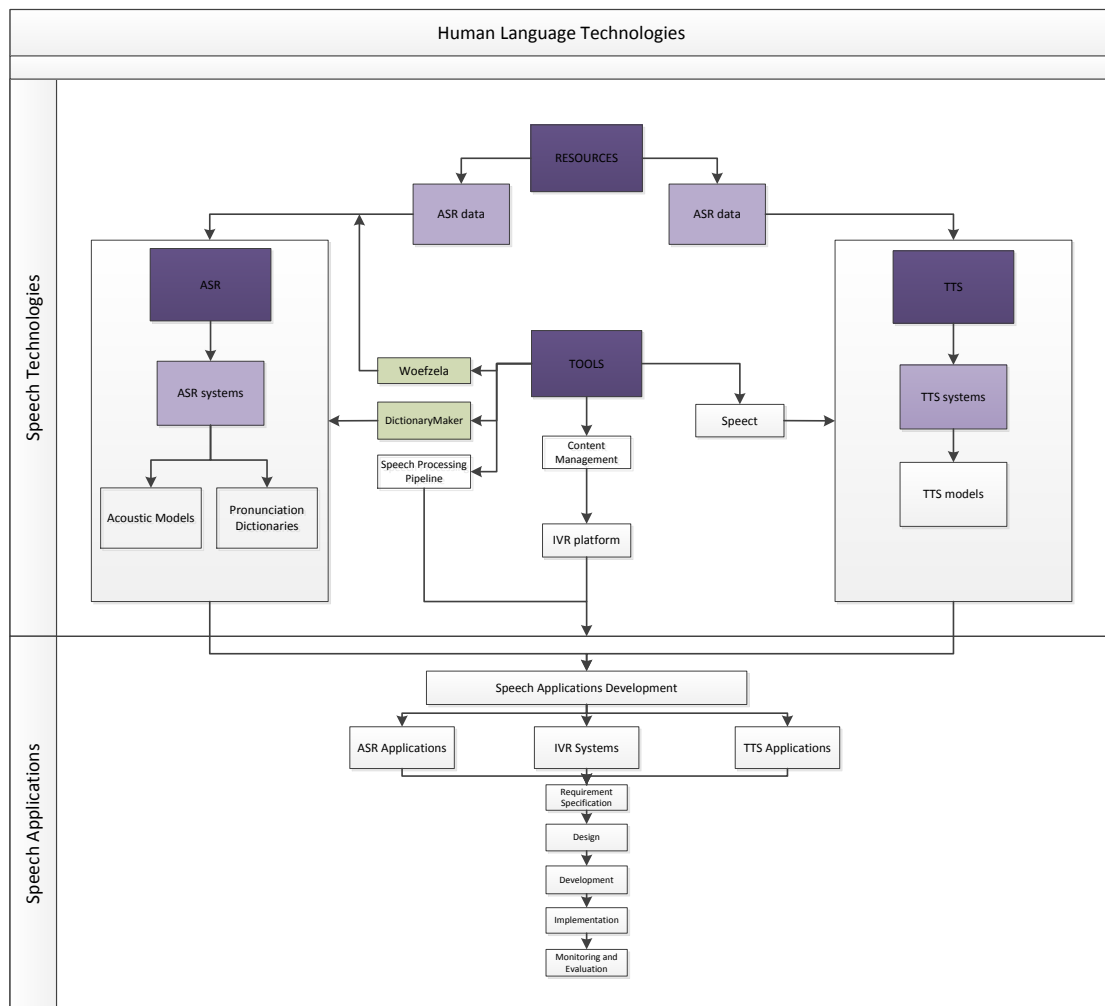


Figure 44: High-level technology overview

15.2 Lwazi II technologies to be transferred

The technologies which are available for transfer, can be categorised into –

- resources;
- systems and software; and
- applications.

Different transfer channels apply to each of these.

15.2.1 Transfer of resources

The ASR data, ASR systems, TTS data and TTS voices were transferred to DAC, with the intention of subsequent transfer to the NCHLT Resource Management Agency (RMA). The RMA will determine the conditions under which the resources will be made available to interested parties, which could include determining appropriate licences, fees and transfer mechanisms. Some resources may have restricted use. For example, in the case of the

Lwazi II project, the TTS voices and application prompts may not be used for commercial purposes as the voice artists did not consent to such use. The following Lwazi II resources were released to DAC:

Table 50: Resource release

Resource	Description	Conditions for release	Release status
ASR Data	The HLT team developed an ASR data collection protocol in order to generate the specifically curated phone-sets, pronunciation dictionaries and speech corpora needed to power ASR in each of the official South African languages. These ASR systems will be released to the RMA (via DAC) in the form of READMEs, some scripts, a set of acoustic models and a pronunciation dictionary which will be licensed.	Respondents did not consent to the commercial use of their voices.	Release for research purposes only.
TTS Data	A voice artist is recruited and speech based on specially prepared text is recorded in studio. These recordings are processed and annotated with phone alignments to form a speech corpus for each language.	Respondents did not consent to the commercial use of their voices.	Release for research purposes only.
TTS Voices	A set of text processing modules and acoustic models for the Speect synthesis system is used to implement TTS in each language.	Respondents did not consent to the commercial use of their voices.	Release for research purposes only.

15.2.2 Transfer of systems and software

The Lwazi I and II systems and software released on a shared platform known as SourceForge (<http://sourceforge.net/>), are indicated in Table 51. SourceForge is a web-based source code repository. It acts as a centralized location for software developers to control and manage free and open source software development and its release. A software release typically enables a user to download the software code as well as readme documents, release notes, user manuals and tutorials for the specific product.

Table 51: Software release

Technology	URL	Status
Telephony system	http://sourceforge.net/projects/lwazi/	Updated in 2011
Speect Version 0.9.5 Version 1.0.0 Version 1.1.0	http://sourceforge.net/projects/speect/	Released in 2010 Released in 2011 Released in 2012
DictionaryMaker Version 2.0	http://sourceforge.net/projects/dictionarymaker/	Released in 2009
DictionaryMaker Version 3.0	http://code.google.com/p/dictionarymaker/	Beta version released in 2012
Speech processing pipeline	http://code.google.com/p/hlt-asr-pipeline/	Alpha version released in 2012

15.2.3 Transfer of applications

The DBE and Afrivet applications consist of various technology building blocks, which need to be integrated in order to render a service, namely –

- the IVR code;
- the voice prompts;
- the content management system code (which is closely integrated with the applications and thus application specific); and
- the telephony platform.

Of these, the telephony platform is the only technology which can be released as a stand-alone offering (see Table 51). The voice prompts can only be released for research purposes, as the voice artists did not consent to the use of their voices for commercial purposes. The rest of the software code (IVR and CMS) is application specific and can at most be released for research purposes.

15.2.4 Licensing

Table 52 summarises the licences under which the Lwazi I and II technologies have been released. The technologies have been released as freely as possible.

Table 52: Licensing

Technology	Lwazi I	Lwazi II
Resources		
Text-to-speech	CC BY-NC	CC BY-NC-SA
Automatic Speech Recognition	CC BY-SA	Transferred to DAC
Pronunciation Dictionaries	CC BY-SA	CC BY-SA
Systems		
ASR Builder	BSD	n/a
DictionaryMaker	BSD	n/a
Lwazi Telephony Platform	BSD	BSD
Speect	MIT	MIT
Speech Processing Pipeline	n/a	MIT
Applications		
IVR: CDW implementation software	BSD	n/a
IVR: DBE implementation software	n/a	n/a ²⁴
IVR: Afrivet implementation software	n/a	n/a
Lwazi Service Voice prompts	Creative Commons (attribution 2.5 share and remix)	n/a

16. Impact

The impact which the project has had can be measured in terms of three distinct types of impact:

- Societal impact (relating more to the services subproject)
- Human capital development (relating more to the technologies subproject)
- Scientific impact (relating to both subprojects).

24

16.1 Societal impact

The main aim of the Lwazi II project was to increase the impact of speech technologies in South Africa. Several activities throughout the project aimed to ensure, understand and enhance the desired impact.

16.1.1 In-depth requirements analysis sessions

The selection of both deployed services (DBE and Afrivet), their subsequent designs and their evaluations were based on the outcome of intensive requirements analysis sessions. A set of detailed templates (see annexure I) guided this process, which involved conducting needs assessments, functional specifications analyses and technical specifications analyses. The requirements analysis processes culminated in the development of project charters for each of the services deployed full-time, which were signed off by the clients. These guided the design and development processes and are now part of the speech application development process of the HLT RG.

16.1.2 Rapid prototyping and testing

The approach taken with the deployed services was to design a prototype call flow, test this with a small subset of users, using the Wizard-of-Oz technique, and refine the call flow before undertaking software development on the IVR. This enabled us to design a call flow which met the needs of the client and the user, and also shortened the time to implementation. Feedback on the prototype releases and the subsequent releases was given in the discussion on each service above.

16.1.3 Understanding the business drivers

An in-depth analysis was made of business drivers and design decisions prevalent in the development of multilingual IVRs in South Africa (Van Huyssteen *et al.*, 2012). A model, consisting of 17 applicable business drivers, was developed. In subsequent work (Calteaux *et al.*, 2012), we applied the model to a voice service case study, the DBE Learner feedback service, and grouped the 17 business drivers into two categories, namely –

1. **primary drivers:** cost savings, increased customer satisfaction, and improved access to information and services; and
2. **secondary drivers:** improved branding, revenue generation, customer retention, customer delight, increased call centre agent morale, increased agent utilization, improved productivity, access to business intelligence for strategic advantage, opportunities for upselling of products, compliance with laws and regulations, political motivation, competitive advantage, response to pain-points, and multi-channel consistency.

We found that **cost saving**, **increased customer satisfaction** and **improved access to services and information** were the primary business drivers for the Learner feedback service. The main design issues we identified for this use case were **language offering**, **persona design** and **input modality**. Future research will consider the model in terms of a commercial voice service offering and determine the prevalent design decisions for such a use case.

16.1.4 Business modelling

In June 2012, a visiting researcher from TNO (the Dutch equivalent of the CSIR) collaborated with the HLT RG on understanding the business model for the Afrivet services. Using the Business Model Canvas methodology (Osterwalder and Pigneur, 2010) and a financial tool, a business model for the Afrivet services was developed. Apart from this, the research visit also led to skills transfer on business modelling for the so-called “bottom of the pyramid”²⁵, by means of several workshops over a period of a week. This work has provided a springboard for improving our understanding of technology transfer and the possibilities for commercialising voice services.

16.2 Human capital development

Table 53 and Table 54 indicate the PhD and Master’s studies respectively completed within the Lwazi II project, as well as those started within the project but not yet completed. Two PhD studies and three Master’s studies were completed, two students resigned and seven studies will continue beyond project closure.

Table 53: PhD studies

Student	Topic of study	University	Status
Jaco Badenhorst	Trajectory modelling with limited speech data	NWU	Ongoing
Thipe Modipa	Automatic recognition of code-switched speech in Sotho-Tswana languages	NWU	Ongoing
Georg Schlünz	A semantic model of affect for improved prosody in text-to-speech synthesis	NWU	Ongoing
Neil Kleynhans	Automatic speech recognition for resource-scarce environments	NWU	Ongoing
Charl van Heerden	Estimating SVM parameters from data	NWU	Completed
Jama Ndwe	Usability engineering of an interactive voice response system in low literacy users of Southern Africa	UCT	Completed

²⁵ The largest but poorest socio-economic group.

Table 54: Master's studies

Student	Topic of study	University	Status
Laurette Pretorius	Developing an enhanced natural language grammar for improved prosody in concept-to-speech synthesis	UNISA	Ongoing
Willem Basson	Morphological decomposition for LVCSR for Afrikaans	NWU	Ongoing
Raymond Molapo	Implementing a distributed approach for speech resource and system development	NWU	Ongoing
Mpho Kgampe	Multilingual proper name pronunciation for speech recognition: the effect of speaker profile	TUT	Ongoing
Nic de Vries	Effective ASR data collection for under-resourced languages	NWU	Completed
Mpho Raborife	Tone labelling algorithm for Sesotho	WITS	Completed
Georg Schlünz	The effects of part-of-speech tagging on text-to-speech synthesis for resource-scarce languages	NWU	Completed

16.3 Scientific impact

16.3.1 Overview of scientific publications

A total of 35 publications emanated from the Lwazi II project. Of these, 13 were published in local conference proceedings, nine in international conference proceedings, and one in an international journal. Table 55 provides an overview of the publications per type, target audience and field of study.

Table 55: Publications overview

Type of publication	Target audience	ASR	TTS	Applications
Conference papers	Local	10	4	
	International	5	4	9
Journal papers	Local	1		
	International			1
Book chapters				1

Abstracts of the publications, per field of study, appear in sections 16.3.2–16.3.4 below.

Table 56 presents an overview of the publications relating to ASR over the life of the project. There was a steady stream of publications on ASR-related topics in all three years of the project and good exposure to both local and international audiences.

Table 56: ASR publications

Title	Author's	Type	Year	Conference/journal name	Abstract no.
Realizations of a single high tone in Northern Sotho	S. Zerbian, E. Barnard	Local journal	2009	Southern African Linguistic and Applied Language Studies	1
Combining regression and classification methods for improving automatic speaker age recognition	C. van Heerden, E. Barnard, M.H. Davel, C. van der Walt, E. van Dyk, M. Feld, C. Muller	International conference	2010	International Conference on Acoustics, Speech and Signal Processing	2
Pooling ASR data for closely related languages	C. van Heerden, N. Kleynhans, E. Barnard, M.Davel	International conference	2010	Workshop on Spoken Languages Technologies for Under-Resourced Languages	3
Learning rules and categorization networks for language standardization	G.B. van Huyssteen, M.H. Davel	International conference	2010	Workshop on Extracting and Using Constructions in Computational Linguistics (co-located with HLT-NAACL)	4
Verifying pronunciation dictionaries using conflict analysis	M.H. Davel, F. de Wet	International conference	2010	11 th Annual Conference of the International Speech Communication Association (Interspeech)	5
Analysing co-articulation using frame-based feature trajectories	J. Badenhorst, M.H. Davel, E. Barnard	Local conference	2010	21 st Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	6
Consistency of cross-lingual pronunciation of South African personal names	M. Kgampe, M.H. Davel	Local conference	2010	21 st Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	7

Title	Author's	Type	Year	Conference/Journal name	Abstract no.
Pronunciation modelling of foreign words for Sepedi ASR	T. Modipa, M.H. Davel	Local conference	2010	21 st Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	8
Towards understanding the influence of SVM hyperparameters	C.J. van Heerden, E. Barnard	Local conference	2010	21 st Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	9
Acoustic modelling of Sepedi affricates for ASR	T. Modipa, M. Davel, F. de Wet	Local conference	2010	20 th Annual Research Conference of the South African Institute for Computer Scientists and Information Technologists (SAICSIT)	10
Efficient harvesting of internet audio for resource-scarce ASR	M.H. Davel, C. van Heerden, N. Kleynhans, E. Barnard	International conference	2011	12 th Annual Conference on the International Speech Communication Association (Interspeech)	11
Trajectory behaviour at different phonemic context sizes	J. Badenhorst, M.H. Davel, E. Barnard	Local conference	2011	22 nd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	12
The predictability of name pronunciation errors in four South African languages	M. Kgampe, M. Davel	Local conference	2011	22 nd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	13
Improved transition models for cepstral trajectories	J. Badenhorst, M. Davel, E. Barnard	Local conference	2012	23 rd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	14
Acoustic model optimisation for a call routing system	N. Kleynhans, R. Molapo, F. de Wet	Local conference	2012	23 rd Annual Symposium of the Pattern Recognition Association of South Africa	15

Title	Author's	Type	Year	Conference/journal name	Abstract no.
				(PRASA)	
Context-dependent modelling of English vowels in Sepedi code-switched speech	T. Modipa, M. Davel, F. de Wet	Local conference	2012	23 rd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	16

Table 57 presents an overview of the publications on TTS-related topics emanating from the Lwazi II project. Again, both local and international conferences were targeted in order to disseminate TTS research findings to a broad audience.

Table 57: TTS publications

Title	Author's	Type	Year	Conference/journal name	Abstract no.
From tone to pitch in Sepedi	E. Barnard, S. Zerbian	International conference	2010	Workshop on Spoken Languages Technologies for Under-Resourced Languages	17
An intonation model for TTS in Sepedi	D.R. van Niekerk, E. Barnard	International conference	2010	11 th Annual Conference of the International Speech Communication Association (Interspeech)	18
Part-of-speech effects on text-to-speech	G.I. Schlünz, E. Barnard, G.B. van Huyssteen	Local conference	2010	21 st Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	19
An African solution for an African problem: A step towards perfection	M. Raborife, S. Ewert, S. Zerbian	Local conference	2010	21 st Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	20

Title	Author's	Type	Year	Conference/journal name	Abstract no.
Experiments in rapid development of accurate phonetic alignments for TTS in Afrikaans	D. van Niekerk	Local conference	2011	22 nd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	21
Developing a corpus to verify the performance of a tone labelling algorithm	M. Raborife, S. Ewert, S. Zerbian	Local conference	2011	22 nd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)	22
Tone realisation in a Yoruba speech recognition corpus	D.R. van Niekerk, E. Barnard	International conference	2012	Workshop on Spoken Languages Technologies for Under-Resourced Languages	23
Introducing the speect speech synthesis platform	J.A. Louw, D.R. van Niekerk, G.I. Schlünz	International conference	2010	The Blizzard Challenge 2010 Workshop	24

Table 58 summarises the publications relating to speech-based applications which emanated from the Lwazi II project. It is significant to note that all of the publications took place in international journals and conference proceedings, placing the Lwazi II applications on the international speech application landscape.

Table 58: Speech applications publications

Title	Author's	Type	Year	Conference/journal name	Abstract no.
Morphological analysis: A method for selecting ICT applications in South African government service delivery	M. Plauche, A. de Waal, A. Sharma, T. Gumede	International journal	2010	Information Technologies and International Development (ITID)	25

Title	Author's	Type	Year	Conference/journal name	Abstract no.
Speech technology for information access: a South African case study	E. Barnard, M.H. Davel, G.B. van Huyssteen	International conference	2010	AAAI Symposium on AI for Development	26
Comparing two developmental applications of speech technology	A. Sharma Grover, E. Barnard	International conference	2011	Conference on Human Language Technologies for Development (HLTD)	27
The Lwazi community communication service: design and piloting of a voice-based information service	A. Sharma Grover, E. Barnard	International conference	2011	20 th International World Wide Web Conference (WWW)	28
Mobile user experience for voice services: A theoretical framework	A. Botha, A. Sharma, K. Calteaux, M. Herselman, E. Barnard	International conference	2012	3 rd International Conference on Mobile Communication for Development (M4D)	29
A voice service for user feedback on school meals	A. Sharma Grover, K. Calteaux, E. Barnard, G.B. van Huyssteen	International conference	2012	2 nd Annual Symposium on Computing for Development (ACM DEV)	30
Aspects of a legal framework for language resource management	A. Sharma Grover, A. Nieman, G.B. van Huyssteen, J.C. Roux	International conference	2012	8th International Conference on Language Resources and Evaluation (LREC)	31
New product development processes for ICT – for development projects	B. McAlister, J. Steyn	International conference	2012	Portland International Centre for Management of Engineering and Technology Conference (PICMET)	32

Title	Author's	Type	Year	Conference/journal name	Abstract no.
Business drivers and design choices for multilingual IVR's: A government service delivery case study	K. Calteaux, A. Sharma Grover, G.B. van Huyssteen	International conference	2012	Workshop on Spoken Languages Technologies for Under-Resourced Languages (SLTU)	33
Voice user interface design for emerging multilingual markets	G.B. van Huyssteen, A. Sharma Grover and K. Calteaux,	Book chapter	2012	Language Sciences and Language Technology in Africa	34
Towards an information ecosystem for animal disease surveillance using voice services	A. Sharma Grover, G.B. van Huyssteen, K. Calteaux	International conference	2013	3 rd Annual Symposium on Computing for Development (ACM DEV)	35

16.3.2 Abstracts – ASR publications

1.

S. Zerbian and E. Barnard. “Realisations of a single high tone in Northern Sotho”, *Southern African Linguistics and Applied Language Studies*, Vol 27(4), pp 357-379, 2009.²⁶

This article reports on a production study that investigates the realisation of a single high tone in the verbal constituent in Northern Sotho, a Bantu language spoken in South Africa. The parameters of variation investigated are based on existing descriptive and theoretical literature and relate to numbers of syllables in the verb stem, morphosyntactic constituency and verb-internal morphological boundaries. The results are interpreted as both phonological and phonetic influences on high tone realisation in this language: phonologically, a high-toned object concord causes a peak shift one syllable to the right. Phonetically, the study shows that the F0 peak associated with a high tone is not necessarily reached within the syllable carrying the high tone but only later (peak delay) depending on the segmental make-up of the tone-bearing syllable and its position within the utterance. The segmental make-up of the tone-bearing syllable leads to systematic surface variation in tone realisation. By collecting controlled acoustic data on tone realisation, this study provides a ground for cross-dialectal comparison of Southern Bantu tone.

<http://www.ajol.info/index.php/salas/article/view/49497>

2.

C. van Heerden, E. Barnard, M. Davel, C. van der Walt, E. van Dyk, M. Feld, C. Muller. “Combining regression and classification methods for improving automatic speaker age recognition”, In *Proceedings of the 35th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp 5174-5177, Dallas, Texas, March 2010.

We present a novel approach to automatic speaker age classification, which combines regression and classification to achieve competitive classification accuracy on telephone speech. Support vector machine regression is used to generate finer age estimates, which are combined with the posterior probabilities of well-trained discriminative gender classifiers to predict both the age and gender of a speaker. The authors show that this combination performs better than direct 7-class classifiers. The regressors and classifiers are trained using long term features such as pitch and formants, as well as short term (frame-based) features derived from MAP adaption of GMMs that were trained on MFCCs.

<http://www.meraka.csir.co.za/pubs/barnard10combiningregression.pdf>

²⁶ Appeared in 2010 and not reported on in Lwazi I report.

3.

C. van Heerden, N. Kleynhans, E. Barnard and M. Davel. “Pooling ASR data for closely related languages”, In Proceedings of the 2nd International Workshop on Spoken Languages Technologies for Under-resourced Languages (SLTU), pp 17-23, Penang, Malaysia, May 2010.

The authors describe several experiments that were conducted to assess the viability of data pooling as a means to improve speech-recognition performance for under-resourced languages. Two groups of closely related languages from the Southern Bantu language family were studied, and our tests involved phoneme recognition on telephone speech using standard tied-triphone Hidden Markov models. Approximately 6 to 11 hours of speech from around 170 speakers was available for training in each language. The authors find that useful improvements in recognition accuracy can be achieved when pooling data from languages that are highly similar, with two hours of data from a closely related language being approximately equivalent to one hour of data from the target language in the best case. However, the benefit decreases rapidly as languages become slightly more distant, and is also expected to decrease when larger corpora are available. Their results suggest that similarities in triphone frequencies are the most accurate predictor of the performance of language pooling in the conditions studied here.

<http://www.meraka.org.za/pubs/vanHeerden10poolingasr.pdf>

4.

G.B. van Huyssteen and M.H. Davel. “Learning rules and categorization networks for language standardization”, In Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL HLT), pp 39-46, Los Angeles, California, June 2010.

In this research, we use machine learning techniques to provide solutions for descriptive linguists in the domain of language standardization. With regard to the personal name construction in Afrikaans, we perform function learning from word pairs using the Default & Refine algorithm. We demonstrate how the extracted rules can be used to identify irregularities in previously standardized constructions and to predict new forms of unseen words. In addition, we define a generic, automated process that allows us to extract constructional schemas and present these visually as categorization networks, similar to what is often being used in Cognitive Grammar. We conclude that computational modeling of constructions can contribute to new descriptive linguistic insights, and to practical language solutions.

<http://www.aclweb.org/anthology-new/W/W10/W10-0806.pdf>

4.

M.H. Davel and F. de Wet. “Verifying pronunciation dictionaries using conflict analysis”, In Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech), pp 1898-1901, Makuhari, Japan, September 2010.

The authors describe a new language-independent technique for automatically identifying errors in an electronic pronunciation dictionary by analyzing the source of conflicting patterns directly. They evaluate the effectiveness of the technique in two ways: they perform a controlled experiment using artificially corrupted data (allowing us to measure precision and recall exactly); and then apply the technique to a real-world pronunciation dictionary, demonstrating its effectiveness in practice. They also introduce a new freely available pronunciation resource (the RCRL Afrikaans Pronunciation Dictionary), the largest such dictionary that currently exists.

<http://www.meraka.org.za/pubs/davel10VerifyingPronDicts.pdf>

6.

J. Badenhorst, M.H. Davel and E. Barnard. “Analysing co-articulation using frame-based feature trajectories”, In Proceedings of the 21st Annual Symposium of the Pattern Recognition Association of South Africa, pp 13-18, Stellenbosch, South Africa, November 2010.

The authors investigate several approaches aimed at a more detailed understanding of co-articulation in spoken utterances. They find that the Euclidean difference between instantaneous frame-based feature values and the mean values of these features are most useful for these purposes, and that low-order polynomials are able to model the between-phone transitions accurately. Examples of typical transitions are presented, and shown to give useful insights on the measurable effects of co-articulation.

<http://www.prasa.org/index.php/2012-03-07-10-55-15>

7.

M. Kgampe and M.H. Davel. “Consistency of Cross-lingual pronunciation of South African personal names”, In Proceedings of the 21st Annual Symposium of the Pattern Recognition Association of South Africa, pp 123-127, Stellenbosch, South Africa, November 2010.

We investigate the consistency with which speakers with different language profiles are able to pronounce personal names from Afrikaans, English, Setswana and isiZulu. We gather data in a controlled research study and analyse cross-lingual pronunciation effects. We find that speakers with a similar primary language tend to agree on the ‘correct’ pronunciation of a name originating from their own language community, and that the ability of speakers from other language communities to approximate this pronunciation is highly dependent on the speaker-word language pair. We also find that there are systematic ways in which names

are 'mis-pronounced' by different language communities: understanding such systematicity could be important when extending electronic pronunciation dictionaries (used in spoken dialogue systems) with the most important variants that occur in practice, in order to increase the accuracy of name recognition.

<http://www.prasa.org/index.php/2012-03-07-10-55-15>

8.

T. Modipa and M.H. Davel. "Pronunciation Modelling of foreign words for Sepedi ASR", In Proceedings of the 21st Annual Symposium of the Pattern Recognition Association of South Africa, pp 185-189, Stellenbosch, South Africa, November 2010.

This study focuses on the effective pronunciation modelling of words from different languages encountered during the development of a Sepedi automatic speech recognition (ASR) system. While the speech corpus used for training the ASR system consists mostly of Sepedi utterances, many words from English (and other South African languages) are embedded within the Sepedi sentences. In order to model these words effectively, different approaches to pronunciation dictionary development are investigated, specifically: (1) using language-specific letter-to-sound rules to predict the pronunciation of each word (based on the language of the word) and mapping foreign phonemes to Sepedi phonemes using linguistically motivated mappings, (2) experimenting with data-driven foreign-to-Sepedi phonemes using linguistically motivated mappings, and (3) using Sepedi letter-to-sound to predict the pronunciation of all words irrespective of language. We find that the data-driven phoneme mappings are more accurate than the initial linguistically motivated mappings evaluated, and (with a slight margin) obtain our best result using Sepedi letter-to-sound rules across all words in the speech corpus.

<http://www.prasa.org/index.php/2012-03-07-10-55-15>

9.

C.J. van Heerden and E. Barnard. "Towards understanding the influence of SVM hyperparameters", In Proceedings of the 21st Annual Symposium of the Pattern Recognition Association of South Africa, pp 281-286, Stellenbosch, South Africa, November 2010.

In this article, the authors investigate the relationship between SVM hyperparameters for linear and RBF kernels and classification accuracy. The process of finding SVM hyperparameters usually involves a gridsearch, which is both time-consuming and resource-intensive. On large datasets, 10-fold cross-validation grid searches can become intractable without supercomputers or high performance computing clusters. They present theoretical and empirical arguments as to how SVM hyperparameters scale with N, the amount of learning data. By using these arguments, the authors present a simple algorithm for finding

approximate hyperparameters on a reduced dataset, followed by a focused line search on the full dataset. Using this algorithm gives comparable results to performing a grid search on complete datasets.

<http://www.prasa.org/index.php/2012-03-07-10-55-15>

10.

T. Modipa, M.Davel and F. de Wet. “Acoustic modelling of Sepedi affricates for ASR”, In Proceedings of the South African Institute for Computer Scientists and Information Technologists (SAICSIT), pp 394-398, Bela-Bela, South Africa, October 2010.

Automatic speech recognition (ASR) systems are increasingly being developed for under-resourced languages, especially for use in multilingual spoken dialogue systems. We investigate different approaches to the acoustic modelling of Sepedi affricates for ASR. We determine that it is possible to model various of these complex consonants as a sequence of much simpler sounds. This approach reduces the Sepedi phoneme inventory from 45 to 32, resulting in simpler dictionary development and transcription processes, as well as more accurate acoustic modelling.

<http://www.meraka.org.za/pubs/Modipa10Acoustic.pdf>

11.

M.H. Davel, C. van Heerden, N. Kleynhans and E. Barnard. “Efficient harvesting of internet audio for resource-scarce ASR”, In Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech), pp 3153-5156, Florence, Italy, August 2011.

Spoken recordings that have been transcribed for human reading (e.g. as captions for audio visual material, or to provide alternative modes of access to recordings) are widely available in many languages. Such recordings and transcriptions have proven to be a valuable source of ASR data in well-resourced languages, but have not been exploited to a significant extent in under-resourced languages or dialects. Techniques used to harvest such data typically assume the availability of a fairly accurate ASR system, which is generally not available when working with resource-scarce languages. In this work, the authors define a process whereby an ASR corpus is bootstrapped using unmatched ASR models in conjunction with speech and approximate transcriptions sourced from the Internet. They introduce a new segmentation technique based on the use of a phone-internal garbage model, and demonstrate how this technique (combined with limited filtering) can be used to develop a large, high-quality corpus in an under resourced dialect with minimal effort.

http://researchspace.csisr.co.za/dspace/bitstream/10204/5769/1/Davel_2011.pdf

12.

J. Badenhorst, M.H. Davel and E. Barnard. “Trajectory behaviour at different phonemic context sizes”, In Proceedings of the 22nd Annual Symposium of the Pattern Recognition Association of South Africa, pp 1-6, Vanderbijlpark, South Africa, November 2011.

The authors propose a piecewise-linear model for the temporal trajectories of Mel Frequency Cepstral Coefficients during phone transitions. As with conventional Hidden Markov Models, the parameters of the model can be estimated for different phonemic context sizes, but their model allows for an intuitive understanding of the impact of context size. They find that the most detailed models, predictably, match the coefficient tracks best - but when data scarcity forces them to use less detailed models, different types of context modelling (clustered triphones versus biphones) have complimentary behaviours. The authors discuss how this complementarity may be useful for data-efficient ASR

<http://www.prasa.org/index.php/2012-03-07-10-55-15>

13.

M. Kgampe and M.H. Davel. “The predictability of name pronunciation errors in four South African languages”, In Proceedings of the 22nd Annual Symposium of the Pattern Recognition Association of South Africa, pp 85-90, Vanderbijlpark, South Africa, November 2011.

Personal names are often pronounced in very different ways depending on the language background of the speaker. We seek to determine whether some of these pronunciations ‘errors’ are systematic and if so, in which ways. Specifically, we analyze some of the typical errors made by speakers from four South African languages (Setswana, English, isiZulu) when producing names from the same four languages. We compare these results with the pronunciations generated by four language-specific grapheme-to-phoneme (G2P) predictors trained on generic words from the four languages. We find that the G2P predictors are able to predict at least some of the typical errors humans make and, in fact, that these errors are slightly more predictable than the correct pronunciations themselves.

<http://www.prasa.org/index.php/2012-03-07-10-55-15>

14.

J. Badenhorst, M.H. Davel and E. Barnard. “Improved transition models for cepstral trajectories”, In Proceedings of the 23rd Annual Symposium of the Pattern Recognition Association of South Africa, pp 157-164, Pretoria, South Africa, November 2012.

We improve on a piece-wise linear model of the trajectories of Mel Frequency Cepstral Coefficients, which are commonly used as features in Automatic Speech Recognition. For

this purpose, we have created a very clean single-speaker corpus, which is ideal for the investigation of contextual effects on cepstral trajectories. We show that modelling improvements, such as continuity constraints on parameter values and more flexible transition models, systematically improve the robustness of our trajectory models. However, the parameter estimates remain unexpectedly variable within triphone contexts, suggesting interesting challenges for further exploration.

<http://www.prasa.org/index.php/2012-03-07-10-55-15>

15.

N. Kleynhans, R. Molapo and F. de Wet. “Acoustic model optimisation for a call routing system”, In Proceedings of the 23rd Annual Symposium of the Pattern Recognition Association of South Africa, pp 165-172, Pretoria, South Africa, November 2012.

The paper presents work aimed at optimising acoustic models for the Auto-Secretary call routing system. To develop the optimised acoustic models: (1) an appropriate phone set was selected and used to create a pronunciation dictionary, (2) various cepstral normalization techniques were investigated, (3) three South African corpora and multiple training data combinations were used to train the acoustic models, and, (4) model-space transformations were applied. Using an independent testing corpus, which contained proper names and South African language names, a named-language recognition accuracy of 95.11 % and proper name recognition accuracy of 93.31% were obtained.

<http://www.prasa.org/index.php/2012-03-07-10-55-15>

16.

T.I. Modipa, M.H. Davel and F. de Wet. “Context-dependent modelling of English vowels in Sepedi code-switched speech”, In Proceedings of the 23rd Annual Symposium of the Pattern Recognition Association of South Africa, pp 173-178, Pretoria, South Africa, November 2012.

When modelling code-switched speech (utterances that contain a mixture of languages), the embedded language often contains phones not found in the matrix language. These are typically dealt with by either extending the phone set or mapping each phone to a matrix language counterpart. We use acoustic log likelihoods to assist us in identifying the optimal mapping strategy at a context dependent level (that is, at triphone, rather than monophone level) and obtain new insights in the way English/Sepedi code-switched vowels are produced.

<http://www.prasa.org/index.php/2012-03-07-10-55-15>

16.3.3 Abstracts – TTS publications

17.

E. Barnard and S. Zerbian. “From tone to pitch in Sepedi”, In Proceedings of the 2nd International Workshop on Spoken Languages Technologies for Under-resourced Languages (SLTU), pp 29-34, Penang, Malaysia, May 2010.

The authors investigate the acoustic realization of tone in continuous utterances in Sepedi (a language in the Southern Bantu family). Human labelers marked each of the 271 syllables in a 15-sentence corpus produced by a single speaker as "high" or "low". Automatic pitch extraction was then used to estimate the fundamental frequencies of the voiced segments of each of these syllables. Statistical analysis of the resulting pitch contours confirms that the mean pitch frequencies of the syllabic nuclei serve as the primary indicator of tone, with the relative frequencies of successive syllables being the most relevant measure. Our analysis also suggests that additional factors may play a role in the production and perception of tone.

<http://www.meraka.org.za/pubs/barnard10tonetopitch.pdf>

18.

D.R. van Niekerk and E. Barnard. “An intonation model for TTS in Sepedi”, In Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech), pp 2182-2185, Makuhari, Japan, September 2010.

The authors present an initial investigation into the acoustic realisation of tone in continuous utterances in Sepedi (a language in the Southern Bantu family). An analytic model for the generation of appropriate pitch contours given an utterance with linguistic tone specification is presented and evaluated. By comparing the model output to speech data from a small tone-marked corpus we conclude that the initial implementation presented here is capable of generating pitch contours exhibiting some realistic properties and identify a number of aspects that require further attention. Lastly, we present some initial perceptual results when integrating the proposed model into a Hidden Markov Model-based speech synthesis system.

http://researchspace.csir.co.za/dspace/bitstream/10204/4661/1/van%20Niekerk_2010.pdf

G.I. Schlünz, E. Barnard and G.B. van Huyssteen. “Part-of-speech effects on text-to-speech synthesis”, In Proceedings of the 21st Annual Symposium of the Pattern Recognition Association of South Africa, pp 257-262, Stellenbosch, South Africa, November 2010.

One of the goals of text-to-speech (TTS) systems is to produce natural-sounding synthesised speech. Towards this end various natural language processing (NLP) tasks are performed to model the prosodic aspects of the TTS voice. One of the fundamental NLP tasks being used is the part-of-speech (POS) tagging of the words in the text. This paper investigates the effects of POS information on the naturalness of a hidden markov model (HMM) based TTS voice when additional resources are not available to aid in the modelling of prosody. It is found that, when a minimal feature set is used for the HMM context labels, the addition of POS tags does improve the naturalness of the voice. However, the same effect can be accomplished by including segmental counting and positional information instead of the POS tags.

<http://www.prasa.org/proceedings/2010/prasa2010-43.pdf>

M. Raborife, S. Ewert and S. Zerbian. “An African solution for an African problem: A step towards perfection”, In Proceedings of the 21st Annual Symposium of the Pattern Recognition Association of South Africa, pp 225-230, Stellenbosch, South Africa, November 2010.

This article reports on a study that involves improving a tone label prediction algorithm. Tone is an important prosodic feature for Bantu languages since these languages use it to distinguish meaning. Studies have shown that text-to-speech systems need detailed prosodic models of a language in order to sound natural to native speakers of the language. Thus, text-to-speech systems developed for Bantu languages need to have tone implemented in them. In order to implement tone, it is necessary for a tone modeling algorithm to receive as input the tone labels of the syllables of a word. This allows the algorithm to predict the appropriate intonation of the word. Our study is concerned with improving an algorithm that predicts tone marks of the syllables of a word. This algorithm uses three Sesotho tonal rules to predict the tone labels on the syllables of polysyllabic verb stems. We improve the algorithm by implementing another tonal rule and extending the application of the tonal rules to other parts of speech.

<http://www.prasa.org/proceedings/2010/prasa2010-38.pdf>

21.

D.R. van Niekerk. “Experiments in rapid development of accurate phonetic alignments for TTS in Afrikaans”, In Proceedings of the 22st Annual Symposium of the Pattern Recognition Association of South Africa, pp 144-149, Vanderbijlpark, South Africa, November 2011.

The quality of corpus-based text-to-speech (TTS) systems depends on the accuracy of phonetic annotation (alignments) which directly influences the process of acoustic modelling. In this paper the author discusses the rapid development of accurate alignments for new languages in an under-resourced context based on an investigation of automatically obtained alignments of an Afrikaans speech corpus. It is shown that certain classes of inaccuracies can be effectively detected by acoustic analyses.

http://researchspace.csir.co.za/dspace/bitstream/10204/5694/1/VanNiekerk_2011.pdf

22.

M. Raborife, S. Zerbian and S.Ewert. “Developing a corpus to verify the performance of a tone labelling algorithm”, In Proceedings of the 22st Annual Symposium of the Pattern Recognition Association of South Africa, pp 126-131, Vanderbijlpark, South Africa, November 2011.

The authors report on a study that involved the development of a corpus used to verify the performance of two tone labelling algorithms, with one algorithm being an improvement on the other. These algorithms were developed for speech synthesis purposes with the aim of improving the perceived naturalness as well as the intelligibility of the speech produced by the synthesizer. The corpus used to test the algorithms consisted of 45 Sesotho sentences specifically chosen to represent the contexts that are necessary for the application of the algorithms. Statistical methods were used to prove the reliability of their transcripts.

http://researchspace.csir.co.za/dspace/bitstream/10204/5657/1/Raborife_2011.pdf

23.

D.R. van Niekerk and E. Barnard. “Tone realisation in a Yoruba speech recognition corpus”, In Proceedings of the 3rd International Workshop on Spoken Languages Technologies for Under-resourced Languages (SLTU), pp 54-59, Cape Town, South Africa, May 2012.

The authors investigate the acoustic realisation of tone in short continuous utterances in Yoruba. Fundamental frequency contours are extracted for automatically aligned syllables from a speech corpus of 33 speakers collected for speech recognition development. Extracted contours are processed and analysed statistically to describe acoustic properties in different tonal contexts. The authors demonstrate how features useful for tone recognition

or synthesis can be successfully extracted from a corpus of this nature and confirm some previously described phenomena in this context.

<http://www.mica.edu.vn/sltu2012/files/proceedings/11.pdf>

24.

J.A. Louw, D.R. van Niekerk and G.I. Schlünz, “Introducing the Speect speech synthesis platform,” in The Blizzard Challenge 2010 Workshop, Kansai Science City, Japan, September, 2010.²⁷

We introduce a new open source speech synthesis engine and related set of tools: Speect is designed to be a portable and flexible synthesis engine, equally relevant as a research platform and runtime synthesis system in multilingual environments. In this paper we document our approach to the rapid development of British English voices for the 2010 Blizzard Challenge using this platform and resources.

http://festvox.org/blizzard/bc2010/MERAKA_Blizzard2010.pdf

16.3.4 Abstracts – Speech applications

25.

M. Plauche, A. de Waal, A. Sharma Grover and T. Gumede. “Morphological analysis: A method for selecting ICT applications in South African government service delivery”, *Information Technologies and International Development (ITID)*, Vol 6 (Spring), pp 1-20, 2010.

Successful ICT projects depend on complex, interrelated sociological and technical factors for which there are no standard theoretical framework for prediction or analysis. Morphological analysis is a problem-solving method for defining, linking, and evaluating problem spaces that are inherently non-quantitative. In this article, the authors show how their research team created a telephony impact model using morphological analysis to strategically select a national ICT telephony project for South Africa from several possibilities, based on non-quantitative, socio-technical criteria. The telephony impact model provides a rigorous framework to the diagnostic and planning phases of our action research that is a vast improvement over “best practices” guidelines. They believe that this approach takes a first step toward predictive models and theories for ICT deployment.

<http://itidjournal.org/itid/article/viewArticle/485>

26.

²⁷ Included here as it has bearing on the TTS engine (Speect).

E. Barnard, M.H. Davel and G.B. van Huyssteen. "Speech technology for information access: a South African case study", In Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI) Spring Symposium, pp 8-13, Palo Alto, California, March 2010.

Telephone-based information access has the potential to deliver a significant positive impact in the developing world. We discuss some of the most important issues that must be addressed in order to realize this potential, including matters related to resource development, automatic speech recognition, text-to-text speech systems, and user-interface design. Although our main focus has been on the eleven official languages of South Africa, we believe that many of these same issues will be relevant for the application of speech technology throughout the developing world.

<http://www.meraka.org.za/pubs/barnard10speechtechnology.pdf>

27.

A. Sharma Grover and E. Barnard. "Comparing two developmental applications of speech technology", In Proceedings of the Conference on Human Language Technology for Development, pp 81-86, Alexandria, Egypt, May 2011.

Over the past decade applications of speech technologies for development (ST4D) have shown much potential for enabling information access and service delivery. In this paper the authors review two deployed ST4D services and posit a set of dimensions that pose a conceptual space for the design, development and implementation of ST4D applications. They also reflect on these dimensions based on their experiences with the above-mentioned services and some other well-known projects.

<http://www.hltd.org/pdf/HLTD201114.pdf>

28.

A. Sharma Grover and E. Barnard. "The Lwazi community communication service: Design and piloting of a voice-based information service", In Proceedings of the 20th International World Wide Web Conference (WWW), pp 433-442, Hyderabad, India, April 2011.

The authors present the design, development and pilot process of the Lwazi Community Communication Service (LCCS), a multilingual automated telephone-based information service. The service acts as a communication and dissemination tool that enables managers at local community centres to broadcast information (e.g. health, employment, social grants) to community workers and the communities they serve. The LCCS allows the recipients to obtain up-to-date, relevant information in a timely and efficient manner, overcoming the obstacles of transportation, time and costs incurred in trying to physically obtain information from the community centres. The authors discuss their experiences and fieldwork in piloting

the LCCS at six locations nationally in the eleven official South African languages. They analyze the usage pattern from the pilot call logs and thereafter discuss the implications of these findings for future projects that design similar automated services for serving rural communities in developing world regions.

<http://www.conference.org/www2011/proceeding/companion/p433.pdf>

29.

A. Botha, A. Sharma Grover, K. Calteaux, M. Herselman and E. Barnard. “Mobile user experience for voice services: a theoretical framework”, In Proceedings of the 3rd International Conference on Mobile Communication for Development (M4D), pp 335-350, New Delhi, India, February 2012.²⁸

The purpose of this paper is to provide a “Mobile User Experience Framework for Voice services.” The rapid spread of mobile cellular technology within Africa has made it a prime vehicle for accessing services and content. The challenge remains to provide these services and information on the technology that the user already owns and is proficient in using. To this end voice as service and distribution mechanism for information is, ahead of SMS and USSD, the most ubiquitous channel of access as it is available on all mobile phone handsets. User experience has been linked to the uptake and engagement with technology and services. As such it becomes imperative to acknowledge mobile user experiences for voice services in order to provide an optimal engagement opportunity that would facilitate participation by end users.

http://researchspace.csir.co.za/dspace/bitstream/10204/5645/1/Botha_2012.pdf

30.

A. Sharma Grover, K. Calteaux, E. Barnard and G. van Huyssteen. “A voice service for user feedback on school meals”, In Proceedings of the 2nd Annual Symposium on Computing for Development (ACM DEV), article 13, Atlanta, United States of America, March 2012.

Research using voice-based services as a technology platform for providing information access and services within developing world regions has shown much promise. The results for design and deployment of such voice-based services have varied depending on the application domain, user community and context. In this paper the authors describe their work on developing a voice-based service for obtaining feedback from school children, a previously unexplored user community. Through a user study, focus group discussions and observations of learners’ interaction with multiple design prototype versions, they investigated several factors around input modality preference, language preference,

²⁸ Included here as it has bearing on the user experience evaluations of the Lwazi II services.

performance and overall user experience. Whilst no significant differences were observed for performance across the prototypes, there were strong preferences for speech (input modality) and English (language). Focus group discussions revealed rich information on learner's perceptions around trust, confidentiality and general system usage. They highlight several design changes made and provide further recommendations on designing for this user community.

<http://dl.acm.org/citation.cfm?id=2160619>

31.

A. Sharma Grover, A. Nieman, G.B. van Huyssteen and J.C. Roux. “Aspects of a legal framework for language resource management” In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC), pp 1035-1039, Istanbul, Turkey, May 2012.²⁹

The management of language resources requires several legal aspects to be taken into consideration. In this paper we discuss a number of these aspects which lead towards the formation of a legal framework for a language resources management agency. The legal framework entails examination of the agency's stakeholders and the relationships that exist amongst them, the privacy and intellectual property rights that exist around the language resources offered by the agency, and the external (e.g. laws, acts, policies) and internal legal instruments (e.g. end user licence agreements) required for the agency's operation.

http://www.lrec-conf.org/proceedings/lrec2012/pdf/226_Paper.pdf

32.

B. McAlister and J. Steyn. “New product development processes for ICT-for-development projects”, in Proceedings of the Portland International Center for Management of Engineering and Technology (PICMET) Conference in Technology Management for Emerging Technologies, pp 3537-3548, Vancouver, Canada, July 2012.

The potential applicability of established new product development processes to information and communications technology (ICT)-for-development projects is investigated. The demand for ICT solutions to serve numerous societal information needs in developing regions of the world is increasing rapidly. A number of methods and practices have been used by organizations to develop and deliver such ICT solutions, but a need exists to formalize product development processes for use in the ICT-for-development context. Existing literature on product development in the ICT-for-development context is explored

²⁹ Included as it has bearing on the transfer of the Lwazi II resources to the Resource Management Agency.

to derive a theoretical model that may be suitable for addressing product development process problems encountered in such projects. An ICT-for-development project to disseminate government information to rural communities with limited literacy is evaluated against the derived theoretical model in a case study. The project was carried out by the CSIR (Council for Scientific and Industrial Research), a government agency in South Africa. The presence and positive effect of certain established product development practices is identified, while the absence or unsatisfactory execution of other established practices are assessed for their contribution to decreased levels of product success.

<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6304373>

33.

K. Calteaux, A. Sharma Grover and G.B. van Huyssteen. “Business drivers and design choices for multilingual IVRs: A government service delivery case study”, In Proceedings of the 3rd International Workshop on Spoken Languages Technologies for Under-resourced Languages (SLTU), pp 29-35, Cape Town, South Africa, May 2012.

Multilingual emerging markets hold many opportunities for the application of spoken language technologies, such as interactive voice response (IVR) systems. Designing such systems requires an in-depth understanding of the business drivers and salient design decisions pertaining to these markets. In this paper we analyze the business drivers and design issues for a voice service (the School Meals Line) piloted in the public sector. We find that cost saving, increased customer satisfaction and improved access to services and information are the primary business drivers for this use case. The main design issues we identify for this use case, and discuss, are language offering, persona design and input modality.

http://researchspace.csir.co.za/dspace/bitstream/10204/5884/1/Calteaux_2012.pdf

34.

G.B. van Huyssteen, A. Sharma Grover and K. Calteaux, “Voice user interface design for emerging multilingual markets”, Language Sciences and Language Technology in Africa, pp 291-308.

Multilingual emerging markets hold many opportunities for the application of spoken language technologies, such as automatic speech recognition (ASR) or text-to-speech (TTS) technologies in interactive voice response (IVR) systems. However, designing such systems requires an in-depth understanding of the business drivers and salient design decisions pertaining to these markets.

<http://www.africansunmedia.co.za/SuneShop/ProductDetails/tabid/78/ProductID/309/Default.aspx>

A. Sharma Grover, G.B. van Huyssteen and K. Calteaux, “Towards an information ecosystem for animal disease surveillance using voice services”, In Proceedings of the 3rd Annual Symposium on Computing for Development (ACM DEV), Bangalore, India, January 2013.³⁰

In this paper we introduce a solution for disease surveillance and monitoring in the primary animal health care (PAHC) domain that uses inbound voice-based services and voice- and text-based outbound services for connecting rural veterinarians and livestock owners with a PAHC service provider. We describe our findings from the ongoing pilots, where we found that it is crucial to close the loop between data collection and information dissemination.

<http://dev2013.org/posters/dev13posters-final12.pdf>

³⁰ Included here as it has a direct bearing on the Lwazi II services. Publication details not yet available.

17. Summary

To summarise, we list the project deliverables as they relate to the various contractual milestones. As can be seen, all deliverables for the project have been completed.

Table 59: Lwazi deliverables

Milestone	Planned due date	Expected progress	Current progress
Service development, piloting and deployment			
Additional piloting: pilot service 1 (CDW service)	30 June 2010	100%	100%
Development and piloting: service 2 & 3	31 December 2010	100%	100%
Development and piloting: service 4	30 June 2011	100%	100%
Development and piloting: service 5	31 December 2011	100%	100%
Deployment of services	31 December 2012	100%	100%
Platform development			
Platform: speech processing pipeline analysis & design	30 June 2010	100%	100%
Platform: initial content management system	31 December 2010	100%	100%
Platform: initial speech processing pipeline	31 December 2010	100%	100%
Platform: improved content management system	31 December 2011	100%	100%
Platform: improved speech processing pipeline	31 December 2011	100%	100%
Platform: future infrastructures report ³¹	31 December 2012	100%	100%
Automatic speech recognition (ASR)			
ASR: system optimisation	31 December 2010	100%	100%
ASR: system	31 December 2011	100%	100%

³¹ Attached as annexure K.

Milestone	Planned due date	Expected progress	Current progress
optimisation			
ASR: data collection	31 December 2012	100%	100%
ASR: final system implementation	31 December 2012	100%	100%
Text-to-speech (TTS) Conversion			
TTS: NLP analysis	31 December 2010	100%	100%
TTS: NLP analysis	31 December 2011	100%	100%
TTS: data collection	31 December 2012	100%	100%
TTS: NLP implementation	31 December 2012	100%	100%
Technology evaluation			
Technology evaluation	31 December 2012	100%	100%
Technology release			
Release	31 December 2012	100%	100%

18. References

Badenhorst, J. Davel M.H. and Barnard, E. 2012. Improved transition models for cepstral trajectories. *Proceedings of the 23rd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2012)*, Pretoria, South Africa, November 2012, pp. 157-164.

Badenhorst, J., Davel M.H. and Barnard, E. 2010. Analysing co-articulation using frame-based feature trajectories. *Proceedings of the 21st Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2012)*, Stellenbosch, South Africa, November 2010, pp. 13-18.

Badenhorst, J., Davel M.H. and Barnard, E. 2011. Trajectory behaviours at different phonemic context-sizes. *Proceedings of the 22nd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2012)*, Vanderbijlpark, South Africa, November 2011, pp. 1-6.

Barnard, E. and Zerbian, S. 2010. From tone to pitch in Sepedi. *Proceedings of the 2nd International Workshop on Spoken Languages Technologies for Under-Resourced Languages (SLTU 2010)*, Malaysia, Penang, May 2010, pp. 29-34.

Boersma, P. 2001. *Praat, a system for doing phonetics by computer*. Amsterdam: Glott International, 2001.

Calteaux, K., Grover A.S. and Van Huyssteen, G.B. 2012. Business drivers and design choices for multilingual IVRs: A government service delivery case study. *Proceedings of the 3rd International Workshop on Spoken Languages Technologies for Under-resourced Languages (SLTU 2012)*, Cape Town, South Africa, May 2012, pp. 29-35.

Carnegie Mellon University. 2012. CMU_ARCTIC speech synthesis databases. Available online at: http://festvox.org/cmu_arctic/. Date accessed: 20 December 2012.

CSIR Meraka Institute. 2012. Technical Report: NCHLT Project Annual Management Report. Pretoria: CSIR Meraka Institute.

CSIR Meraka Institute. 2011. Technical Report: NCHLT Project Annual Management Report. Pretoria: CSIR Meraka Institute.

CSIR Meraka Institute. 2009. Technical Report: Lwazi Project Final Report: Development of a Telephone-based Speech-driven Information Service for the South African Government. Pretoria: CSIR Meraka Institute.

Davel, M.H. and De Wet, F. 2010. Verifying pronunciation dictionaries using conflict analysis. *Proceedings of the 11th Annual Conference of the Speech Communication Association (Interspeech 2010)*, Makuhari, Japan, September 2010, pp. 1898-1901.

Diambra, J.F., McClam, T., Fuss, A., Burton, B. and Fudge, D.L. 2009. Using a Focus Group to Analyse Students' Perceptions of a Service-learning Project. *The College Street Journal*, 34(1): 114-122.

Du Plessis, J.A., Gildenhuis J.G. and Moiloa, J.J. 1974. *Tweetalige woordeboek Afrikaans-Suid-Sotho / Bukantswe ya maleme-pedi Sesotho-Seafrikanse*. Cape Town: Via Afrika Beperk, First Edition.

Greeff, M. 2005. Information collecting: Interviewing. In De Vos, A.S., Strydom, H., Fouché, C.B. and Delpont, C.S.L. (eds). *Research at Grass Roots for the Social Services and Human Services Professions (3rd ed.)*. Pretoria: Van Schaik Publishers, pp. 286-313.

Grover, A.S., Calteaux, K., Barnard E. and Van Huyssteen, G. 2012. A voice service for user feedback on school meals. *Proceedings of the 2nd Annual Symposium on Computing for Development (ACM DEV 2012)*, Atlanta, United States of America, March 2012, Article 13.

Grover, A.S., Stewart O. and Lubensky, D. 2009. Designing interactive voice response (IVR) interfaces: localisation for low literacy users. *Proceedings of the 12th IASTED International Conference on Computers and Advanced Technology in Education (CATE)*, St. Thomas, US Virgin Islands, November 2009, Article 673-029.

Kgampe, M. and Davel, M.H. 2010. Consistency of cross-lingual pronunciation of South African personal names. *Proceedings of the 21st Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2010)*, Stellenbosch, Cape Town, November 2010, pp. 123-127.

Kgampe, M. and Davel, M.H. 2011. The predictability of name pronunciation errors in four South African languages. *Proceedings of the 22nd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2011)*, Vanderbijlpark, South Africa, November 2011, pp. 85-90.

Kleynhans, N., Molapo R. and De Wet, F. 2012. Acoustic model optimisation for a call routing system. *Proceedings of the 23rd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2012)*, Pretoria, South Africa, November 2012, pp. 165-172.

Kominek, J., Schultz, T. and Black, A. 2008. Synthesizer Voice Quality of New Languages Calibrated with Mean Mel Cepstral Distortion. *Proceedings of the International Workshop on Spoken Language Technology for Under-Resourced Languages (SLTU)*, 2008.

Leedy, P.D. and Ormrod, J.E. 2005. *Practical Research: Planning and design (8th ed.)*. New Jersey: Pearson Educational International and Prentice Hall.

Louw, J.A., Van Niekerk, D.R. and Schlünz, G.I. 2010. Introducing the Speect speech synthesis platform. *The Blizzard Challenge 2010 Workshop*, Japan, Kansai Science City, September, 2010.

Modipa T. and Davel, M.H. 2010. Pronunciation modelling of foreign words for Sepedi ASR. *Proceedings of the 21st Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2010)*, Stellenbosch, South Africa, November 2010, pp. 185-189.

Modipa, T., Davel M. and De Wet, F. 2010. Acoustic modelling of Sepedi affricates for ASR. *Proceedings of the South African Institute for Computer Scientists and Information Technologists (SAICSIT 2010)*, Bela-Bela, South Africa, October 2010, pp. 394-398.

Modipa, T.I., Davel, M.H. and De Wet, F. 2012. Context-dependent modelling of English vowels in Sepedi code-switched speech. *Proceedings of the 23rd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2012)*, Pretoria, South Africa, November 2012, pp. 173-178.

Osterwalder, A. and Pigneur, Y. 2010. *Business Model Generation*. New Jersey: John Wiley & Sons, Inc.

Raborife, M. 2011. Tone labelling algorithm for Sesotho. Master's dissertation, School of Computer Science, University of the Witwatersrand, Johannesburg.

Raborife, M., Ewert, S. and Zerbian, S. 2010. An African Solution for an African Problem: A step towards perfection. *Proceedings of the 21st Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2010)*, South Africa, Stellenbosch, November 2010, pp. 225-230.

Raborife, M., Zerbian, S. and Ewert, S. 2011. Developing a corpus to verify the performance of a tone labelling algorithm. *Proceedings of the 22nd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2011)*, Vanderbijlpark, South Africa, 2011, pp. 126-131.

Raborife, M., Zerbian, S. and Ewert, S. 2012. Empirical measurements on a Sesotho tone labelling algorithm. *Proceedings of the 3rd International Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU 2012)*, Cape Town, South Africa, May 2012, pp. 133-138.

Schlünz, G.I., Barnard, E. and Van Huyssteen, G.B. 2010. Part-of-speech effects on text-to-speech synthesis. *Proceedings of the 21st Annual Symposium of the Pattern Recognition*

Association of South Africa (PRASA 2010), Stellenbosch, South Africa, November 2010, pp. 257-262.

Van Heerden, C., Kleynhans, N., Barnard E. and Davel, M. 2010. Pooling ASR data for closely related languages. *Proceedings of the 2nd International Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU 2010)*, Penang, Malaysia, May 2010, pp. 17-23.

Van Huyssteen, G.B., Grover, A.S. and Calteaux, K. 2012. Voice user interface design for emerging multilingual markets. In Ndinga-Koumba-Binza, H.S. & Bosch, S.E. (eds). *Language Science and Language Technology in Africa: Festschrift for Justus C. Roux*. Stellenbosch: Sun Media, pp 291-308.

Van Niekerk, D.R. and Barnard, E. 2010. An intonation model for TTS in Sepedi. *Proceedings of the 11th Annual Conference of the Speech Communication Association (Interspeech 2010)*, Makuhari, Japan, September 2010, pp 2182-2185.

Van Niekerk, D.R. and Barnard, E. Predicting utterance pitch targets in Yorùbá for tone realisation in speech synthesis. Submitted for publication.

Van Niekerk, D.R. and Barnard, E. 2012. Tone realisation in a Yorùbá speech recognition corpus. *Proceedings of the 3rd International Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU 2012)*, Cape Town, South Africa, May 2012, pp. 54-59.

Van Niekerk, D.R, Barnard, E. and Schlünz, G.I. 2009. A perceptual evaluation of corpus-based speech synthesis techniques in under-resourced environments. *Proceedings of the 20th Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2009)*, Stellenbosch, South Africa, November/December 2009, pp. 71-75.

Van Niekerk, D.R. 2011. Experiments in rapid development of accurate phonetic alignments for TTS in Afrikaans. *Proceedings of the 22nd Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2011)*, Vanderbijlpark, South Africa, November 2011, pp. 144-149.

Van Niekerk, D.R. and Barnard, E. 2009. Phonetic alignment for speech synthesis in under-resourced languages. *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, Brighton, United Kingdom, September 2009, pp. 880-883.

Van Santen, J.P.H. and Buchsbaum, A.L. 1997. Methods for optimal text selection. *Proceedings of the European Conference on Speech Communication and Technology (Eurospeech 1997)*, Rhodes, Greece, September 1997, pp. 553-556.

Yamagishi, J., Nose, T. Zen, H., Ling, Z., Toda, T. Totuda, K., King, S. and Renals, S. 2009. Robust speaker-adaptive HMM-based text-to-speech synthesis. *IEEE Transactions on Audio, Speech and Language Processing*, 17(6), pp. 1208-1230.

Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T. and Kitamura T. 2001. Mixed excitation for HMM-based speech synthesis. *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech 2001)*, Aalborg, Denmark, September 2001, pp. 2263-2266.

Annexure A – Letter to schools

Dear Principal

Request to visit the Mthombo secondary school for the evaluation of project

Lwazi

The Human Language Technology (HLT) research group at the Council for Scientific and Industrial Research, in collaboration with the National Department of Basic Education is conducting a research project entitled Lwazi II. The project is funded by the National Department of Arts and Culture. It entails building speech application technologies for resource scarce South African languages. To archive this, we need spoken voices. We have therefore built a system, named "*Mama Nandi*" used by learners and school NSNP coordinators to report on NSNP services at school including food served.

Your school is one of 16 across the country where we have piloted project Lwazi. The next step will be to evaluate the usefulness of the system. This entails:

- A discussion with the school NSNP coordinator (not more than an hour)
- A focus group discussion with twenty (20) learners (60-95 minutes)

We would like to visit Mpumalanga, Mthombo Secondary school, on the 17th and 18th of October 2012. On the 17th we would like to hand out the consent forms for learners to take home and have them signed by their parents / guardians. The aim of the focus groups with learners and coordinators is to obtain inputs on their experience of the system to improve it for future use. On the second day (18th October) we will hold a focus group discussion with the learners who managed to bring back signed consent forms from parents. Because this discussion will take between 60 – 95 minutes, we request that we talk to them after school. Light snacks will be served.

Kind regards,

.....
Tebogo Gumede-Reid
Monitoring and Evaluation Team Leader
Human Language Technology Competency Area
CSIR

tgumede@csir.co.za
012 841 2641

Annexure B – Learner consent form

LWAZI DBE NSNP learner feedback service
Learner consent form
Group discussion

Thank you for agreeing to participate in the group discussion for the Lwazi NSNP research project. The aim of the project is to design a telephone system to share and obtain information on the Department of Basic Education's National School Nutrition Programme (NSNP). The project is being carried out by researchers from the Human Language Technology Competence Area of the CSIR.

The Lwazi telephone system has been in use at your school for a few months now. The next step is to have group discussions with people who have used the telephone system so that we can get information about your experience when you used the telephone system. This will assist us to improve the system. We are not testing you!! We are testing the systems that we are developing.

Taking part in the group discussion is your choice. **There will be no penalty if you choose not to participate in the project** or answer certain questions. If you decide not to participate in the project, you will not lose any benefits you would have otherwise been owed. You are free to withdraw participating in the group discussion at any time. **Your choice to withdraw will not affect your relationship with the School and/or the Department of Basic Education.**

All the information that you will share with us about the system **will be kept confidential**. We will **not** use your name or the name of the school in any future reports or publications. The information from the group discussions will only be used anonymously in internal reports or publications. Your identity will not be shared with anyone. The photos and video recordings of the group discussions will only be used in research and project-related publications.

Taking part in the group discussion will **not cost you anything** and **you will not be paid** for taking part either.

If you sign this consent form you agree that:

- All the people introduced to you at the group discussion are your guardians, so you will follow all the instructions they give you.
- You will provide us with honest feedback on your experience of the Lwazi telephone system.

If you have any questions about this project, please contact Tebogo Gumede Reid **012 841 4835**

If you agree with the terms outlined above, please tick the relevant boxes and then sign below.

I agree to take part in the Lwazi research group discussions.	<input type="checkbox"/>
I agree to be photographed and/or video recorded.	<input type="checkbox"/>

Name of child:

Signature of child: Date: 2012

Contact telephone number:

Address:

Annexure C – Parent consent form

**LWAZI DBE NSNP learner feedback reporting
service**
Parent consent form
Group discussion

Thank you for agreeing to let your child take part in the Lwazi research discussion group. The aim of the project is to design a telephone system to share and obtain information on the Department of Basic Education’s National School Nutrition Programme (NSNP). The project is being carried out by researchers from the Human Language Technology Competence Area of the CSIR.

The Lwazi telephone system has been in use at your child’s school for a few months now. The next step is to have group discussions with children who have used the telephone system so that we can get information about how they experience the telephone system. This will assist us to improve the system. We are not testing the learners!! We are testing the systems that we are developing.

If you sign this consent form you agree that your child can –

- attend a discussion group as explained above;
- be available (and on time) at his/her school on the designated date and at the designated time;
- provide us with honest feedback on his/her experience of using the Lwazi telephone system; and
- be photographed and/or video recorded while participating in the group discussion.

Taking part in this pilot project is your and your child’s choice. **There will be no penalty if you choose for your child not to participate in the project** or answer certain questions. If you decide that your child should not participate in the project, you will not lose any benefits you would have otherwise been owed. You are free to withdraw the participation of your child from this pilot project at any time. Your choice to withdraw your child from the discussion will not affect your relationship with the School and/or the Department of Basic Education.

Information obtained during the group discussion **will be kept confidential**. We will not use your child’s name or the name of your child’s school in any reports or publications. Your child’s identity will not be shared with anyone. The photos and video recordings of the group discussions will only be used in research and project-related publications.

Taking part in the group discussion will **not cost you or your child anything** and **neither you nor your child will be paid** for taking part.

If you have any questions about this study, please contact Tebogo Gumede **012 841 4835**.

If you agree with the terms outlined above, please **tick the relevant boxes** and then **sign** below.

I agree that my child may take part in the Lwazi research group discussion.

I agree that my child may be photographed and/or video recorded.

Name of child:

Signature of guardian /parent: Date: 2012

Name of parent:

Contact telephone number:

Address:

**Annexure D – Learner feedback service FGD
interview schedule**

DBE LEARNER FEEDBACK APPLICATION
User Experience Evaluation Focus Groups



Arts and Culture
Basic Education



Focus group themes

- Experience of the piloting (done at schools)
 - What did the Lwazi ambassadors tell you about Mama Nandi?
 - What was your experience of mama Nandi when you called the first time?
 - When you were not sure what to do, did you go to the ambassador for assistance?
- Usage
 - Do you understand what the system is about?
 - What is their experience of the system: actual system experience?
 - Is there any confusion/anything not clear?
 - What language did you choose and why?
 - For those who chose English, why?
- Uptake
 - How often do you usually call?
 - With the frequency of your use, has it been easier to call?
 - What would make you call more? What makes you call less?
 - What happens when you use the system? Environment, system, friends etc.
- Impact
 - Does the voice of the system make you want to tell her more about the food at school?
 - Is the mama Nandi a safe place for you to leave messages about food at school? Do you trust mama Nandi? Why / not?
 - Are you comfortable with mama Nandi system?
- Motivation
 - What can we do to make sure you call more?
 - What other channels do you or your friends use to complain about the food at school or about the happenings at school?

Programme (3 hours)

Ice breaker 35 Min
Play UNO

Remember to assess spontaneity
Tell us about the food you like/dislike to eat at home 25 Min
Tell us about the food at school

BREAK 20 Min

Lwazi questions 2 Hours

Start discussion with the continuation of food they like / dislike

Annexure E – Coordinator reporting service evaluation questionnaire

DBE COORDINATOR REPORTING APPLICATION
District manager and Coordinator questions (Phase 2 and 3)



Arts and Culture
Basic Education



○ **What do you think of the Lwazi service?**

.....
.....
.....

○ **Why have you been using or not using it?**

.....
.....
.....

○ **If you have not been using it, do you have suggestions on how we can improve usability by you and other coordinators in the country?**

.....
.....
.....

Would you recommend this system to someone else?

- **If yes why?**
- **If no why not?**

.....
.....
.....

**Annexure F – Dip tank service evaluation
consent form**



IFOMU LEMVUME NGOKUHLANGANYELA OCWANINGWENI

Ukuhlolwa kwe-Afrivet Training Services dip tank line.

Loluhelelo wakhiwe ngabakwa-CSIR ngokubambisana nabakwa-Afrivet Training Services Training Services ukuze lusetshenziswe ngabalimi abadiphayo. Lolayini uzohlinzeka u-Afrivet Training Services Training Services ngolwazi lokuthi yini eyenzekayo emathangeni ase-Siyaphambili (uzositshela ngezimo zezilwane, inani lezilwane ezidiphiwe, imikhiqizo yokudipha esetshenzisiwe nanoma iziphi izinkinga enihlangabezana nazo) ukuze kwenziwe ngcono imikhiqizo nokudipha.

Uyacelwa ukuthi uhlanganyele kucwaningo olwenziwa ngabakwa-*CSIR Meraka Institute* kanye nabakwa-*Afrivet Training Services Training Services*. Ukhethiwe njengomuntu ongase ahlanganyele kulolu cwaningo ngoba ungumlimi odipha e-Siyaphambili, okuyindawo lopho abakwa-Afrivet Training Services Training Services benza khona iprojekti.

1. INHLOSO YOCWANINGO

Ngokuvivinya loluhlelo, usiza abakwa-CSIR ukuba balwenze ngcono. Lokhu kuzokwenza uzuthi esikhathini esizayo, uhlelo lolu lwandiselwe kuwo wonke amafama ase-Ningizimu Afrika.

2. IZINQUBO

Uma uvolontiya ukuhlanganyela kulolu cwaningo, sizokucela ukuthi wenze okulandelayo:

- Ulalele izaziso zocwaningo lwaloluhlelo,
- Ushayele inombolo oyinikeziwe bese uphendula imibuzo ephathelene nokudipha,
- Ubikele abacwaningi ukuthi uluthola lunjani uhlelo. Ubatshela ukuthi lungashintshwa kuphi ukuze libe ngcono.

3. OKUNGAHLE KUBE UBUCAYI FUTHI KUNGAKUPHATHI KAHLE

Ukuthi imibuzo ekulolu hlelo ingacaci futhi ungayiqondi.

4. BANGASIZAKALA KANJANI ABAHLANGANYELI NOMA/KANYE NOMPHEKATHI?

Amafama adiphayo angenzelwa ngcono uhlelo lokuqashelwa kokudipha nezifo zezilwane.

5. INKOKHELO NGOKUHLANGANYELA

Ayikho

6. IMFIHLO

Ulwazi olutholakale ngokuxhumene nalolu cwaningo futhi olungayamaniswa nawe, luzohlala luyimfihlo kanti luzodalulwa ngemvume yakho kuphela noma uma lidingwa ngomthetho. Onke amafomu emvume nemibiko, azogcinwa endaweni ephaphile lapho kuzofinyelela khona ngabakwa - CSIR/Afrivet Training Services Training Services kuphela.

Nangaphezu kwaloko, uma usishayela ucingo kulolu hlelo, awuphoqekile ukuthi uziveze ukuthi ungubani noma ushiye inombolo yakho. Yonke imininingwane yalolu hlelo edluliselwe kwikhompyutha izobonwa kuphela ngabakwa-CSIR/Afrivet Training Services Training Services.

7. UKUHLANGANYELA NOKUHOXA

Awuphoqelwanga ukuba ube ingxenye yalolucwaningo. Uma uvolontiya ukuba kulolu cwaningo, yazi ukuthi ungahoxa noma nini futhi kungabi namithelela yanoma iluphi uhlobo kulokho. Uma kukhona umbuzo ongathandi ukuwuphendula, yazi futhi ukuthi uvumelekile ukungawuphenduli, udlulele kweminye . Umcwaningi angakukhipha kulolu cwaningo uma kukhona izimo eziphoqayo ukuthi enze njalo.

8. UKWAZI UKUTHI OBANI ABACWANINGI

Uma unemibuzo noma imiphi, noma kukhona okukukhathazayo maqondana nocwaningo, khululeka uthintane no-Tebogo Gumede ku- 012 841 2614.

9. AMALUNGELO ABAHLANGAYELI

Ungayihoxisa noma inini imvume yakho futhi uyeke ukubamba iqhaza ngaphandle kwesijezo. Ngokuhlanganyela kulolu cwaningo, awulahlekelwa lutho ngokomthetho, ngokwamalungelo, nangokwezixazululo. Uma unemibuzo mayelana namalungelo akho njengomuntu othatha iqhaza kulolucwaningo, thintana no-Dr Sandile Ncanana, wase-CSIR REC Secretariat ehhovisi le-Research and Development kulezizinombolo 012 841 4060 noma uthumele i-imeyili ku- R&DEthics@csir.co.za.

Ukusayinda komhlanganyeli kucwaningo noma omele lapho umhlanganyeli enenkinga yokuzibhalela ngokwakhe

Lencazelo engenhla ichazelwe u- _____ (umhlanganyeli) ngu-Tebogo Gumede (umcwaningi) ngesi-Zulu (ulimi) Ngiye nganikwa ithuba lokubuza imibuzo kanti le mibuzo iye yaphendulwa ngeneliseka [mina/yena].

Isivumelwano sokuhlanganyela ocwaningweni

Ngaloku nginikeza imvume ngokuzikhethela yokuhlanganyela kulolu cwaningo. Nginikeziwe ikhophi yaleli fomu.

_____	_____	_____
Igama lomhlanganyeli	Usuku	Sayinda

_____	_____	_____
Igama lommele	Usuku	Sayinda

(Uma ekhona)

Isivumelwano sokuthathwa isithombe nokuqoshwa kanye/noma ukuthathwa i-video ocwaningweni

Ngaloku ngiyavuma ukuthathwa isithombe kanye/noma ukuthathwa i-video kulolu cwaningo. Nginikeziwe ikhophi yaleli fomu.

_____	_____	_____
Igama lomhlanganyeli	Usuku	Sayinda
_____	_____	_____
Igama lommele	Usuku	Sayinda
(Uma ekhona)		

iSignesha yomcwaningi/obuza imibuzo

Mina _____ (*umcwaningi*)
ngiyaqinisa ukuthi ngimchazele yonke imininingwane ebhalwe kulo mbhalo u-
_____ (*umhlanganyeli*) kanye/noma ommele
_____ (*omele umhlanganyeli*). Uye wagqugquzelwa futhi
wanikwa isikhathi esanele ukuba angibuze noma imiphi imibuzo. Le ngxoxo yenziwe
ngesi-**Zulu** (*ulimi olusetshenzisiwe*). Asibanga bikho isidingo somhumushi njengalo
umhlanganyeli ubeluqonda ulimi olusetshenzisiwe ngenhla noma le ngxoxo iye
yahunyushelwa esi-_____
(*ulimi*) ngu-_____ (*umhumushi*).

_____	_____
Sayinda yomcwaningi/obuza imibuzo	Usuku

Annexure G – Dip tank evaluation question schedule

Post Pilot questionnaire for dip tank attendants

Name of dip tank attendant : _____

Name of researcher : _____

Date & time start : _____

Location : _____

User experience

1. Have you used the Afrivet Training Services Dip tank telephone reporting service?

2. What is your opinion of the service/what do you think about it?

3. When did you start using it?

4. How often have you used it?

5. Did you use it each time there was a dipping session?

6. If no, why not **and** what would make you to call the service each time you have a dipping session?

7. Has using the service helped you at all? Probe how it has helped/why it has not helped?

8. Would you recommend the system to someone else and why? (Probe with "Why" at least 5 times)

Usability

9. Was it easy or difficult to use the service? Probe why?

10. Did you know what to answer to each of the questions? Probe why/tell us more?

11. What problems did you have with the system, e.g. were there any questions that you were not able to answer/ any questions that sounded strange or that didn't make sense?

12. Does the telephone system ask all of the relevant questions that you want to report on for a dipping session? Is there anything else you would like to report on, that we can add to the questions?

13. Do you have any other suggestions on how we can improve the service?

14. Is there anything else you would like to tell us about this telephone system?

THANK YOU FOR YOUR PARTICIPATION

Annexure H – Improved density estimation report

LWAZI II: IMPROVED DENSITY ESTIMATION FOR ASR -
FINAL RELEASE

Version 1.0

Prepared for CSIR Meraka Institute

Charl van Heerden, Marelle Davel and Etienne Barnard

Multilingual Speech Technologies

North-West University

Vanderbijlpark

30 October 2012

TABLE OF CONTENTS

SECTION ONE - INTRODUCTION	2
SECTION TWO - ACOUSTIC MODEL DEVELOPMENT AND EVALUATION	3
2.1 Introduction	3
2.2 Background	4
2.3 Materials and method	5
2.4 Manual verification	7
2.5 Results	8
2.5.1 Overall performance	8
2.5.2 Effect of corpus verification	9
2.5.3 Vocabulary independence	9
2.5.4 Effect of phrase length	10
2.5.5 Effect of vocabulary size	10
2.5.6 Speaker-specific effects	12
2.6 Conclusion	14
SECTION THREE - DEVELOPMENT OF ASR TOOLS	18
3.1 Introduction	18
3.2 Recent improvements	18
3.3 Current capabilities	19
SECTION FOUR - CONCLUSION	22
APPENDIX A - DELIVERABLES: README	23
REFERENCES	33

SECTION ONE

INTRODUCTION

We report on the development of speech-recognition systems that are able to perform accurate recognition on medium-vocabulary tasks, as specified in the Lwazi II proposal. We are able to achieve error rates of less than 5% (our design goal) on all eleven official languages, by using training corpora that contain 60–155 hours of speech per language. The majority of the errors stem from words such as abbreviations, foreign words or names, which do not adhere to the standard orthography of the target language and which are difficult to produce by speakers who are not familiar with the specific terms. We also find that recognition accuracy does not depend strongly on the number of occurrences of a term in the training set or the length of the term to be recognised, and that a few problematic speakers are responsible for a disproportionate number of errors.

This report extends and improves on the report prepared for the *Intermediate Release*, as well as the conference paper that was based on that report [1]. Here, we describe the process followed to improve and evaluate the accuracy of the acoustic models for the eleven languages: Afrikaans, English, isiNdebele, isiXhosa, isiZulu, Sepedi, Sesotho, Setswana, Siswati, Tshivenda and Xitsonga. Details with regard to the development and evaluation process are included in Section 2, while Section 3 lists some of the tool development that was required to achieve the above results, as well as the capabilities of the current set of tools. A *README* file describing the deliverables associated with this report is included as an appendix.

SECTION TWO

ACOUSTIC MODEL DEVELOPMENT AND EVALUATION

2.1 INTRODUCTION

The potential of speech-recognition systems to play a meaningful role in developing-world applications is widely understood [2–4], but the development of such systems for the under-resourced languages which are ubiquitous in the developing world is a major obstacle to the fulfilment of this potential. One way to address this issue is to understand how well recognisers can function if limited resources are used in their development. With this in mind, we have recently shown that flexible and accurate small-vocabulary recognisers can be trained with relatively small amounts of speech [5]. However, recent developments in smartphone-based data collection have significantly reduced the complexity of creating sizable speech corpora in under-resourced languages, thus making it feasible to attempt more ambitious recognition tasks in such languages. Such improved recognition is one of the goals of the Lwazi II project. We have therefore investigated the process of developing a set of recognisers for medium-sized vocabularies (around a few hundred words or phrases per grammar) for all of the eleven South African languages.

Our first goal is to investigate how recognisers trained on 60–155 hours of speech per language perform on realistic medium-vocabulary recognition tasks for information-access applications. Such tasks are characterised by moderately high perplexities (we study perplexities from 50 to 200), variable-length recognition terms and a significant prevalence of proper names and words taken from foreign languages. We then investigate how each of these characteristics affects the performance of the trained recognisers, and also characterise the speaker differences that we observe when this task is performed.

In Section 2.2, we summarise some of the relevant work related to speech recognition for under-resourced languages, including a discussion of the developments related to corpus collection that have enabled us to expand our corpora efficiently. Section 2.3 details the corpus that was used for our experiments, as well as the experimental protocol followed. The manual verification performed as part of this protocol is described in more detail in Section 2.4. Section 2.5 contains our experimental results, and in Section 2.6 we discuss the implications of this work.

2.2 BACKGROUND

Since data scarcity is the basic problem for the development of ASR systems in under-resourced languages, most of the research on such systems has involved some form of data sharing, either with a well-resourced “donor” language, or by pooling data from two or more related under-resourced languages. Several methods using such data sharing are summarised in [6]. Although each of those methods – ranging from phoneme mapping for recognition with “pure” donor-language acoustic models to complete retraining with appropriately pooled data – shows some improvement from data sharing, it is fair to say that the amount of target-language speech remains the most important determinant of the recognition accuracy achieved. That is, none of the pooling strategies is able to deliver comparable accuracy with that which is achieved with sufficient training data from the target language.

Against this background, recent developments in ASR data collection tools are extremely encouraging. In particular, Hughes *et al.* [7] described a smartphone-based application, that greatly streamlines the process of ASR data collection by combining the basic components of such collection in a single, easy-to-use and portable package. The application is loaded with a collection of prompts which are to be recorded by speakers in the target population; each respondent is then handed a smartphone, which displays a sequence of prompts to the speaker and records the resulting speech. This application inspired the development of Woefzela [8], an open-source application for Android smartphones, which is specifically geared towards the practicalities of developing-world data collection – thus, it supports the logistics of field workers who may need to perform data collection in rural or otherwise challenging environments, performs some quality control while the collection is taking place, and does not rely on the constant availability of Internet connectivity.

Using these applications and a collection of, say, ten or fifteen relatively inexpensive smartphones, it is a simple matter to collect speech from a few hundred mother-tongue speakers of a target language in a week or less. As a consequence, the basic assumption that developing-world ASR implies small speech corpora is overturned. Of course, many other challenges remain for ASR in under-resourced languages – the creation of suitable text corpora and pronunciation lexicons as well as user-interface issues and the like remain as formidable obstacles to be overcome in the quest for high-impact speech technology for the developing world.

We can distinguish three overlapping categories of applications which could conceivably con-

tribute to such impact [9, 10]:

- Small-vocabulary recognisers, with vocabularies from 2 to around 20 words, are useful for tasks such as menu traversal and the confirmation of choices or preferences. One can think of such systems as touchtone replacements, which is a useful capability in light of the difficulties that many developing-world users experience with touchtone interfaces [4].
- Medium-vocabulary systems, which typically operate with vocabularies up to several hundred words, can also be used for selecting items from longer lists (for example to choose destinations in a travel application, or to specify a specific location for a service delivery application) or to search for information in limited domains.
- Systems with practically unrestricted vocabularies represent the state of the art in speech recognition; these are used in applications such as voice search, dictation, navigation and numerous others.

During the Lwazi I project, we showed that useable small-vocabulary systems (i.e. with speaker-independent accuracies in the order of 95 % or better) can be developed for under-resourced languages with around 5 – 10 hours of orthographically transcribed speech in the target language. The main goal of the current contribution is to investigate whether similar performance is achievable for medium-vocabulary tasks, given the substantially larger corpora that are rendered feasible with tools such as the application of Hughes *et al.* and Woefzela.

2.3 MATERIALS AND METHOD

Our investigation involves all eleven official languages of South Africa, as shown in Table 2.1. Amongst these, the South African dialect of English (SAE) is the only well-resourced language (and even SAE does not have the benefit of large publicly-available dialect-specific speech corpora or pronunciation lexicons). We employed preliminary versions of speech corpora in those languages that are being developed at the Meraka Institute; statistics of these corpora are also provided in Table 2.1. For each language, its language code is listed: this is used as a shorthand when referring to the various languages later in the report.

Since the corpora all consist of short prompts (mostly 1 – 5 words in length) read by variable numbers of speakers, we constructed our test grammars by randomly selecting 200 prompts from each language as the target utterances. This allows us to investigate the influence of factors such as training-vocabulary dependence and the length of the terms to be recognised. (Due to slightly variable collection protocols employed in the different languages, we had to limit the number of prompts that were spoken by only one or two speakers in some languages. In those languages, we ordered the prompts by their occurrence frequencies in the corpus, and modified our procedure to ensure approximately uniform coverage of the various occurrence frequencies.) For our investigation

Table 2.1: Summary of corpora used in speech-recognition experiments. 8 speakers were held out for future evaluations while the rest of the corpus was used for experiments by performing 4-fold cross-validation.

Language	code	# speakers	# utts.	vocab.	#hours
Afrikaans	afr	190	99 620	3 835	82
English	eng	214	111 374	7 619	80
isiNdebele	ndb	209	90 297	17 666	112
isiZulu	zul	166	82 552	2 793	95
Sesotho	sot	199	106 550	2 660	88
Sepedi	nso	113	66 959	12 581	64
Siswati	ssw	227	123 667	15 332	155
Setswana	tsn	108	65 725	6 212	66
Xitsonga	tso	205	121 456	8 897	143
Tshivenda	ven	194	119 614	9 478	141
isiXhosa	xho	110	76 244	30 046	95

into the effect of variable perplexity on recognition accuracy, random subsets of these 200 target utterances were drawn.

Our protocol does not include a manual verification process of all the recordings – hence, reading and other errors are known to be present in both the training and test sets. In a related DAC-funded project¹ we are developing tools to detect such errors automatically [11, 12], but the present work does not utilise any of those tools. Hence, our error-rate estimates are pessimistic; we return to this matter in Section 2.4 and the conclusion below. (A small subset of recordings have been manually verified – this will be discussed in more detail in Section 2.4.)

The data in each language was split into four partitions with no speaker overlap – each partition thus contained approximately 12 – 15 hours of speech, from approximately 50 speakers. These partitions were then used in cross validation: we repeatedly trained acoustic models on three of the partitions; recognition accuracy was then measured on the remaining portion in each case.

A standard 3-state left to right HMM architecture is used to model context-dependent triphones in each language. As acoustic features, 39-dimensional Mel Frequency Cepstral Coefficients are used: 13 static coefficients with cepstral mean normalisation applied, 13 delta and 13 double delta coefficients. Triphones are tied at the state level using decision tree clustering, and each tied-state triphone is estimated with 8 Gaussian mixtures per state. Semi-tied transforms are also employed throughout. In order to develop these models, we have significantly expanded our training tools; these extensions are summarised in the Appendix at the end of this report.

During speech recognition, all possible phrases in the set of phrases to be recognised are considered equal, and a recognition grammar is constructed that allows the best matching phrase to be selected from the set of possible phrases. This process is repeated across all languages and a cross

¹The National Centre for HLT (NCHLT) speech corpus project

all folds, allowing a detailed investigation of the factors that influence recognition accuracy.

2.4 MANUAL VERIFICATION

Manually verifying all of the utterances and corresponding transcriptions in the corpora used for estimating the acoustic models would be a very resource-intensive and expensive task: together, there are 1 064 058 utterances. At even optimistic utterance verification times of around 10s per utterance, this task would still take a single human approximately 3 000 hours of full-time verification. Such long verification times are clearly not feasible for the general development of ASR corpora in under-resourced languages. For this reason, we are developing tools for automatically verifying the quality of our transcriptions [11, 12]. These tools have not been deployed on the corpora used for estimating our acoustic models; hence a subset of the utterances was manually verified for the purposes of evaluating our acoustic models. The manual verification was performed in two batches:

- All utterances for which the decoder could not find a path (even with pruning switched off completely), were manually verified. Manual verification confirmed that all of these utterances were “empty” (that is, they contained no speech at all, even though their durations were non-zero). These utterances were removed from our corpora.
- The four languages with the highest error rates were selected for manual verification. Utterances from a subset of speakers who had error rates of 20% and more were selected for manual verification. In addition, 50 other utterances were selected randomly in order to ensure that the process was not biased.

Software was developed to allow a human to rapidly listen to a list of audio files, while viewing the corresponding transcription. This software also implements the following protocol, allowing the human to classify an utterance as:

- *0* Empty: an audio file has no speech, but may contain music, background noise or silence.
- *1* Correct: the audio matches the transcription perfectly. Reasonable pronunciation variants are allowed, as well as minor pronunciation errors.
- *2* Partially correct: the audio and the transcription corresponds to some degree but contains obvious errors. This is a rather broad category and includes utterances containing repetitions, hesitations, reading errors, words that are cut at the beginning or end, missing words or erroneously added words: if at least a third of the transcription matched the audio, it was considered partially correct.
- *3* Incorrect: the audio does not match the transcription at all. (For example, a query directed at the data collector, instead of a prompt being read.)

- *r* Replay: It is often necessary to listen to an utterance several times.

The category defined for ‘partially correct’ was broad on purpose: it is non-trivial to determine what the types of errors are that can be considered typical of users of a spoken dialogue, and that a speech recognition system should be able to handle. We therefore assumed that an ASR system should be able to handle all partially correct utterances: in this way we are assured that the reported error rates are an accurate estimate or an under-estimate of the performance achieved.

The four languages with the highest error rates were isiNdebele, isiXhosa, Siswati and Tshivenda. Two human verifiers were tasked to listen to the same set of utterances and score them individually. Where the two transcribers disagreed during individual evaluation, the utterances were reviewed and a consensus opinion obtained. Only utterances that both transcribers agree are either empty (0) or incorrect (3) were removed from the test corpus.

Table 2.2 shows the listener agreement (a) over all categories and (b) over the usable/unusable categories only. (For the latter analysis, all categories 1 or 2 are considered usable, all categories 0 or 3 are considered not.) Most of the inter-transcriber disagreement is with regard to categories 1 and 2: when considering whether a prompt is usable or not, inter-transcriber agreement was at least 97.6% for all four languages.

Table 2.2: Human verification results.

Language	# utterances	% agreement	
		all	usability
Siswati	167	91.62	98.13
Tshivenda	78	93.62	97.60
isiNdebele	107	89.72	100.00
isiXhosa	183	92.90	99.45

2.5 RESULTS

In this section, we describe the performance achieved by each of the newly developed recognisers. We first discuss overall performance (section 2.5.1), before analysing a number of important characteristics of the recognisers, namely: the effect of corpus verification, the vocabulary independence of the recognisers, the effect of phrase length, and the implications of increasing or decreasing the vocabulary size. We conclude with an analysis of some speaker-specific effects.

2.5.1 OVERALL PERFORMANCE

Fig. 2.1 shows the error rates that were achieved for the 200-term task across the eleven languages, as well as the accuracies achieved on various subsets of the data. We see that the fundamental goal of 5% error rate on generic terms was achieved for all languages, with error rates on those terms ranging between 1.1% for English to 4.5% for Sepedi. On the other categories of terms, however, performance

was highly variable – for isiZulu, English, Afrikaans, Sesotho and Siswati good performance was achieved across all classes, whereas isiNdebele, Sepedi, isiXhosa (and to a lesser extent Setswana, Xitsonga and Tshivenda) performance on all the non-standard items was much worse. These differences stem from at least two causes. On the one hand, the sophistication of resources (such as pronunciation dictionaries and pre-processing rules to place text in canonical form) differs significantly across the various languages. Another factor is the difference in recording protocols followed for the various languages: these early versions of the NCHLT corpus have significantly different repetition frequencies of the various prompts by different speakers. We return to these factors in Section 2.6 below.

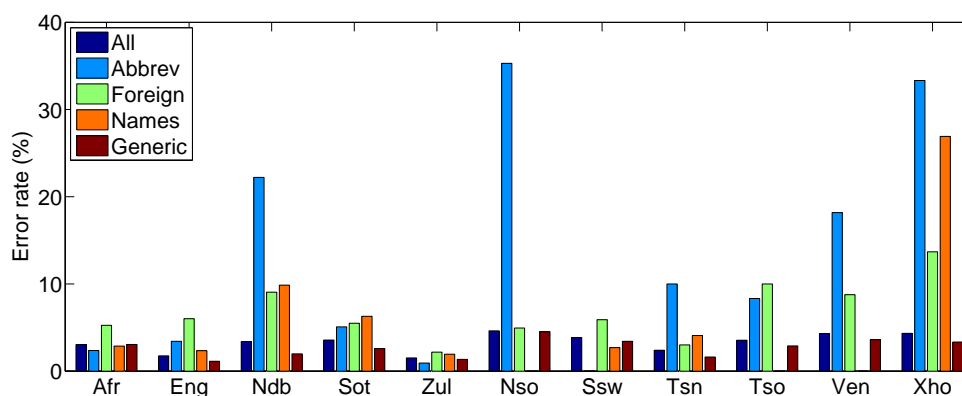


Figure 2.1: Error rates achieved when performing 200-term recognition in several languages. “Abbrev”, “Foreign” and “Names” refer to error rates on subsets of the terms containing one or more words in those categories, whereas “generic” refers to error rates achieved on terms that do not contain words in any of the above categories.

2.5.2 EFFECT OF CORPUS VERIFICATION

As discussed in section 2.4, all unusable utterances were removed from the corpus prior to analysis. This set consisted of all utterances that failed to decode (and were subsequently manually verified to be empty) as well as the utterances labelled as 0 (empty) or 3 (incorrect) during manual verification for four of the eleven languages. The results before and after corpus verification are shown in Table 2.3.

2.5.3 VOCABULARY INDEPENDENCE

We investigate the vocabulary independence of our recognisers, by studying how recognition accuracy of each term depends on the number of occurrences of that term in the training set. Thus, we keep track of the number of instances of each term in a given training fold, and tally the resulting test-fold performance of that term against this training-set count. This process is repeated across all folds. The

Table 2.3: Summary of manual verification results. The number of files which did not decode are indicated for all eleven languages, while only four languages (isiNdebele, isiXhosa, Siswati and Tshivenda) warranted further manual verification. The error rates before and after manual verification are also shown.

Language	# failed decodes	Error before	Error after
Afrikaans	5	3.11	3.04
English	0	1.76	1.76
isiNdebele	0	5.60	3.39
isiZulu	14	1.66	1.52
Sesotho	18	3.83	3.57
Sepedi	2	4.80	4.62
Siswati	21	8.21	3.85
Setswana	0	2.40	2.40
Xitsonga	8	4.08	3.55
Tshivenda	1	5.08	4.31
isiXhosa	22	9.00	4.33

resulting trends of error rate against the number of training occurrences are shown in Fig. 2.2 and 2.3. As before, there seem to be significant differences between the different languages – however, there are no strong training-count dependencies in any of the languages, suggesting that the accuracies achieved are, to a large extent, vocabulary independent.

2.5.4 EFFECT OF PHRASE LENGTH

Fig. 2.4 and 2.5 summarise our findings on another important issue related to the performance of our system, namely the relationship between the accuracy achieved and the length of the term to be recognised. One would generally expect longer terms to be recognised more accurately, since such terms are more distinct from one another. For English, Setswana, Tshivenda and perhaps Sesotho there is evidence of such a trend; however, those trends are generally fairly weak, suggesting that utterance length is not a strong determinant of confusability within the parameters of our investigation.

2.5.5 EFFECT OF VOCABULARY SIZE

We have also studied how the vocabulary size (and thus, in our design, perplexity) influences recognition accuracy. To this end, we have selected subsets with, respectively, 50 and 100 terms from our original 200-term grammar, and performed the same recognition experiments as above with those target grammars for Afrikaans, English, isiNdebele, isiZulu and Sesotho. Fig. 2.6 shows that the dependency of error rate on perplexity is notable in all languages, but that the magnitude of this dependency again differs somewhat across languages. The weak trend for Afrikaans is easy to explain, given that most of the errors that occurred in the Afrikaans test were caused by a small number of speakers (2 or 3) who had various problems using our recording interface – these problems were so

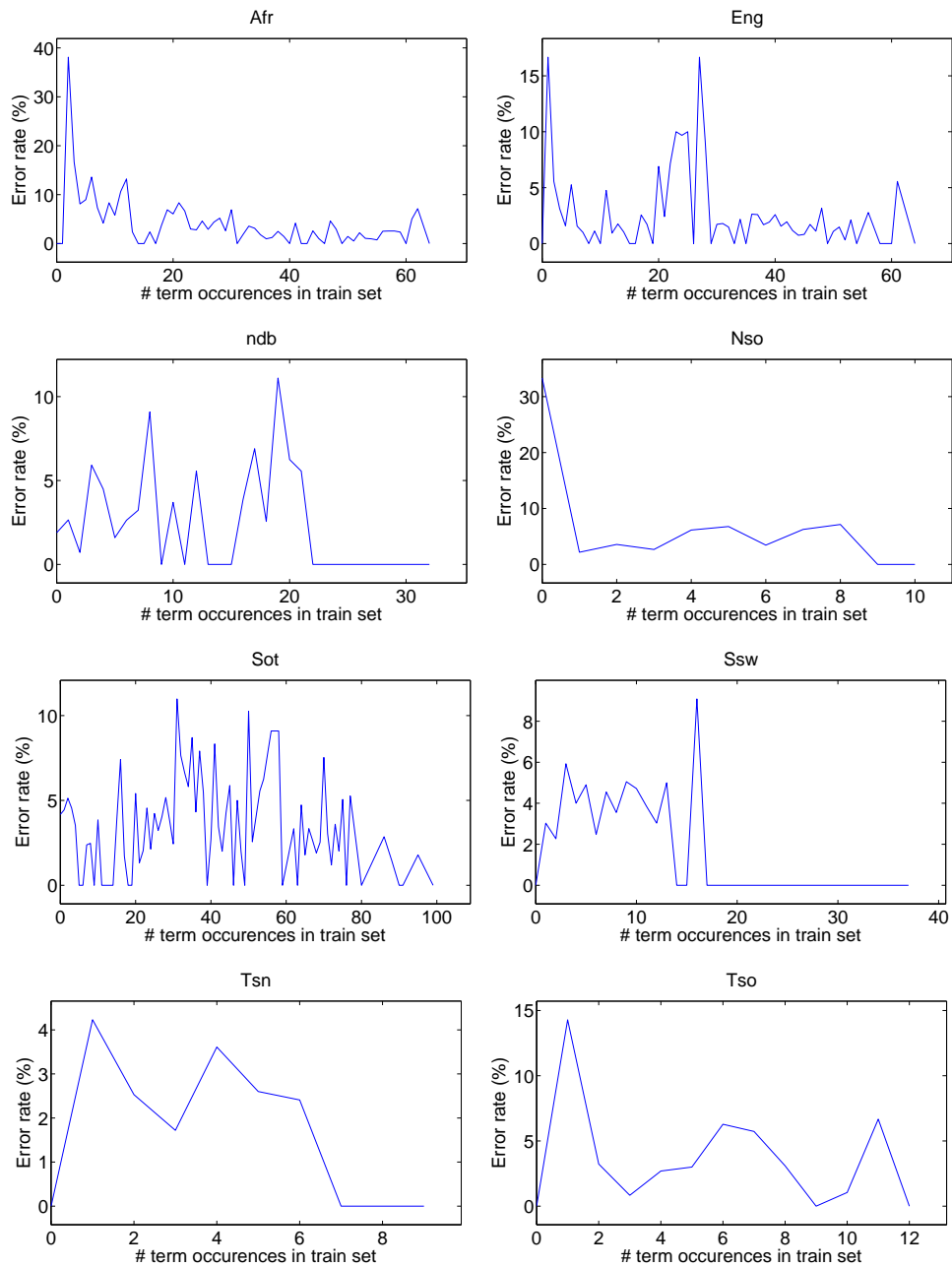


Figure 2.2: Error rates as a function of the number of occurrences of a particular test term in the training set.

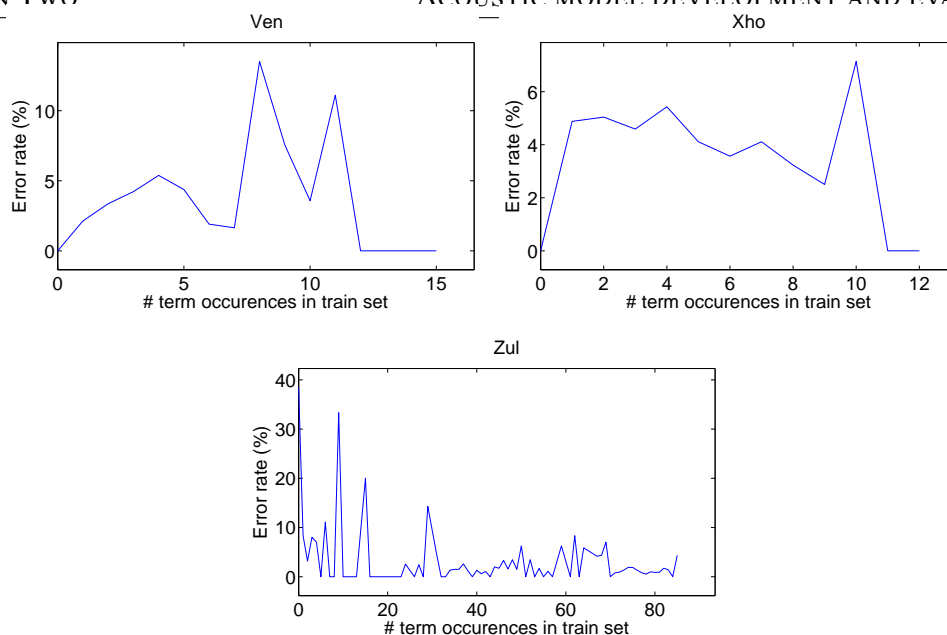


Figure 2.3: Error rates as a function of the number of occurrences of a particular test term in the training set (*continued*).

severe as to cause comparable error rates at all perplexities. The relatively steep curve for isiNdebele does not have such a clear-cut explanation, and deserves further study.

2.5.6 SPEAKER-SPECIFIC EFFECTS

Finally, the error rates for the individual speakers are summarised in Fig. 2.7 and 2.8. In each language, the speakers have been ordered from highest to lowest error rate, and speakers who have contributed fewer than 5 test utterances have been excluded. We see similar trends across all languages, namely one or two speakers with extremely high error rates, followed by an intermediate group whose error rates are in the range 5% to 15%, and a long tail of speakers for whom recognition is perfect or nearly so. The small set of speakers with very high error rates produced poor recordings for a variety of reasons:

- A few speakers had difficulty with the recording hardware, apparently covering the recording microphone while doing the recordings. The resulting audio files are significantly distorted (the high-frequency components are suppressed), but are generally still judged as intelligible by human listeners.
- Others had great difficulty in reading, often producing hesitant speech, partial repetitions and incorrect renderings of the prompts provided. Hence these should more properly be thought of as “reading errors” rather than “recognition errors”.
- A small number of speakers were obviously distracted during reading; their recordings are different from those in the previous category in that the readings are generally fluent, but in-

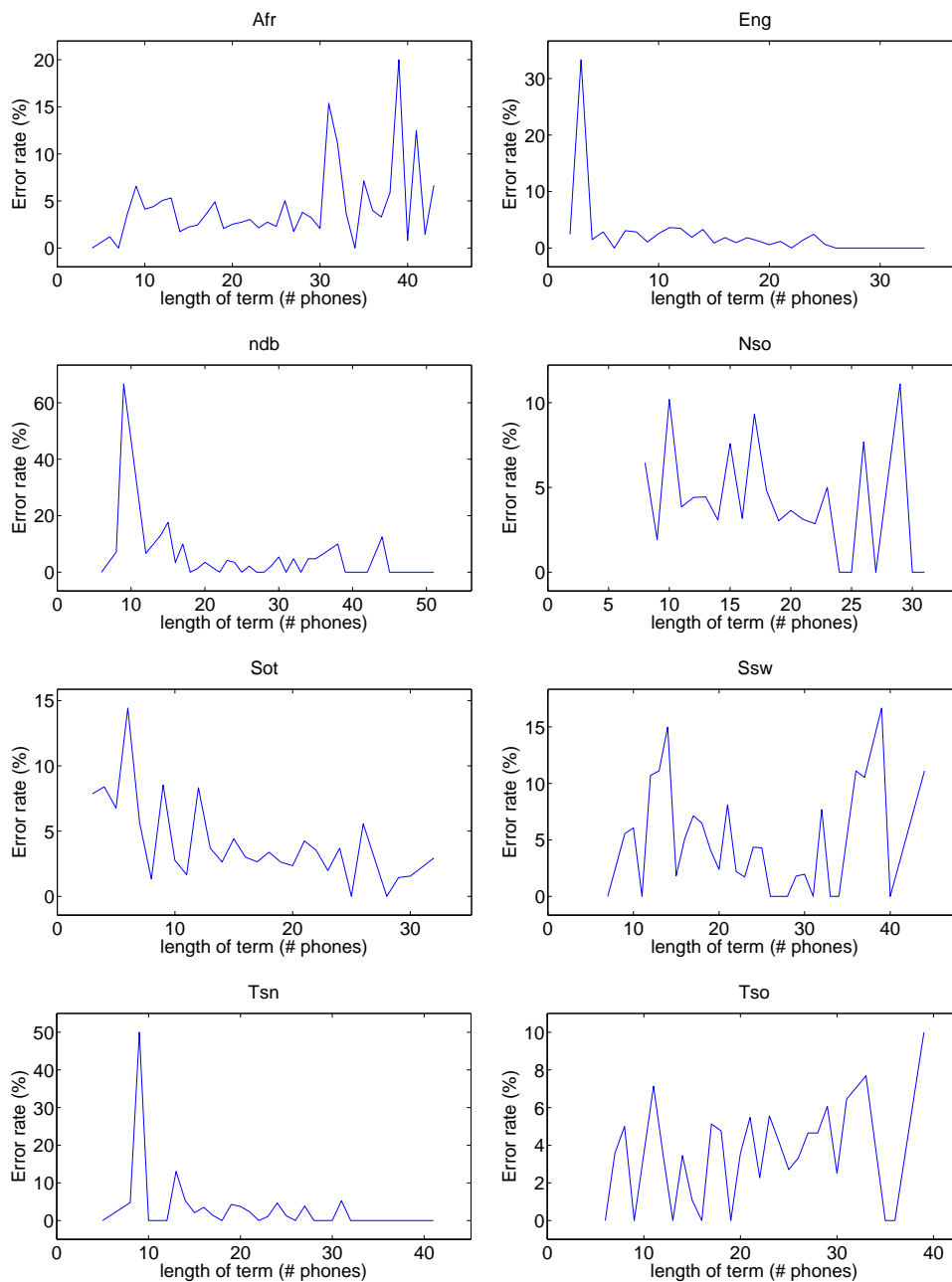


Figure 2.4: Error rates as a function of the number of phones in the term to be recognised (for words with multiple pronunciations, we use the most common pronunciation).

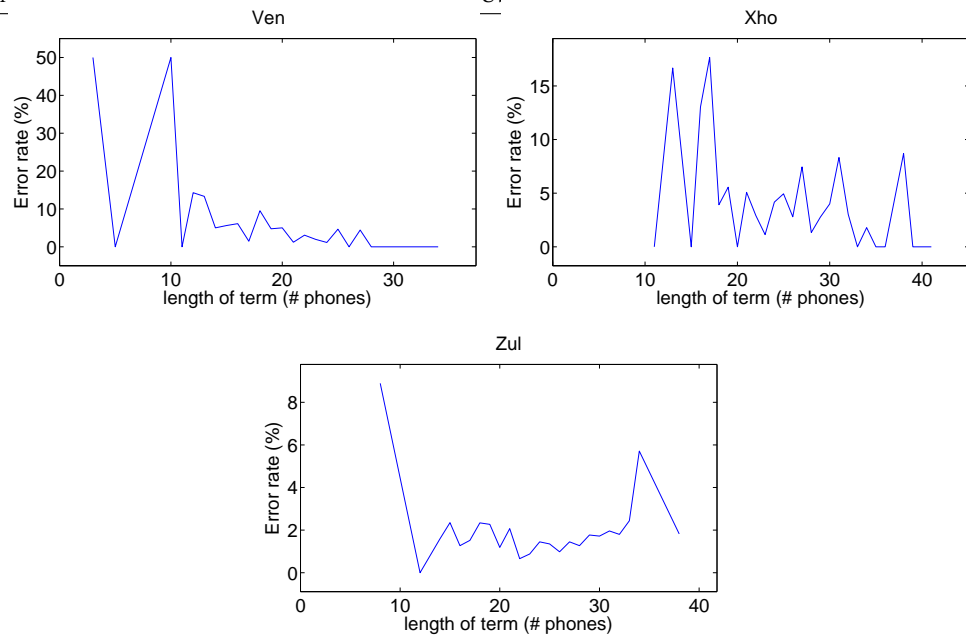


Figure 2.5: Error rates as a function of the number of phones in the term to be recognised (for words with multiple pronunciations, we use the most common pronunciation) (*continued*).

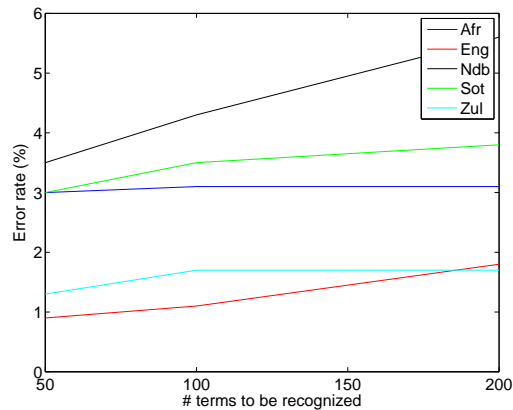


Figure 2.6: Error rate as a function of the number of terms to be recognised.

accurate, containing extraneous speech. Prompts from these speakers are often cut off, as the recording application is not operated successfully by the speaker.

2.6 CONCLUSION

We have shown that useful recognition accuracies can be achieved on medium-vocabulary recognition tasks in low-resource languages using corpora collected with the new generation of smartphone-based data-collection tools. Although details for the languages differ in various ways, a number of common themes emerge:

- The recognition accuracies for words and terms that do not correspond to the standard orthog-

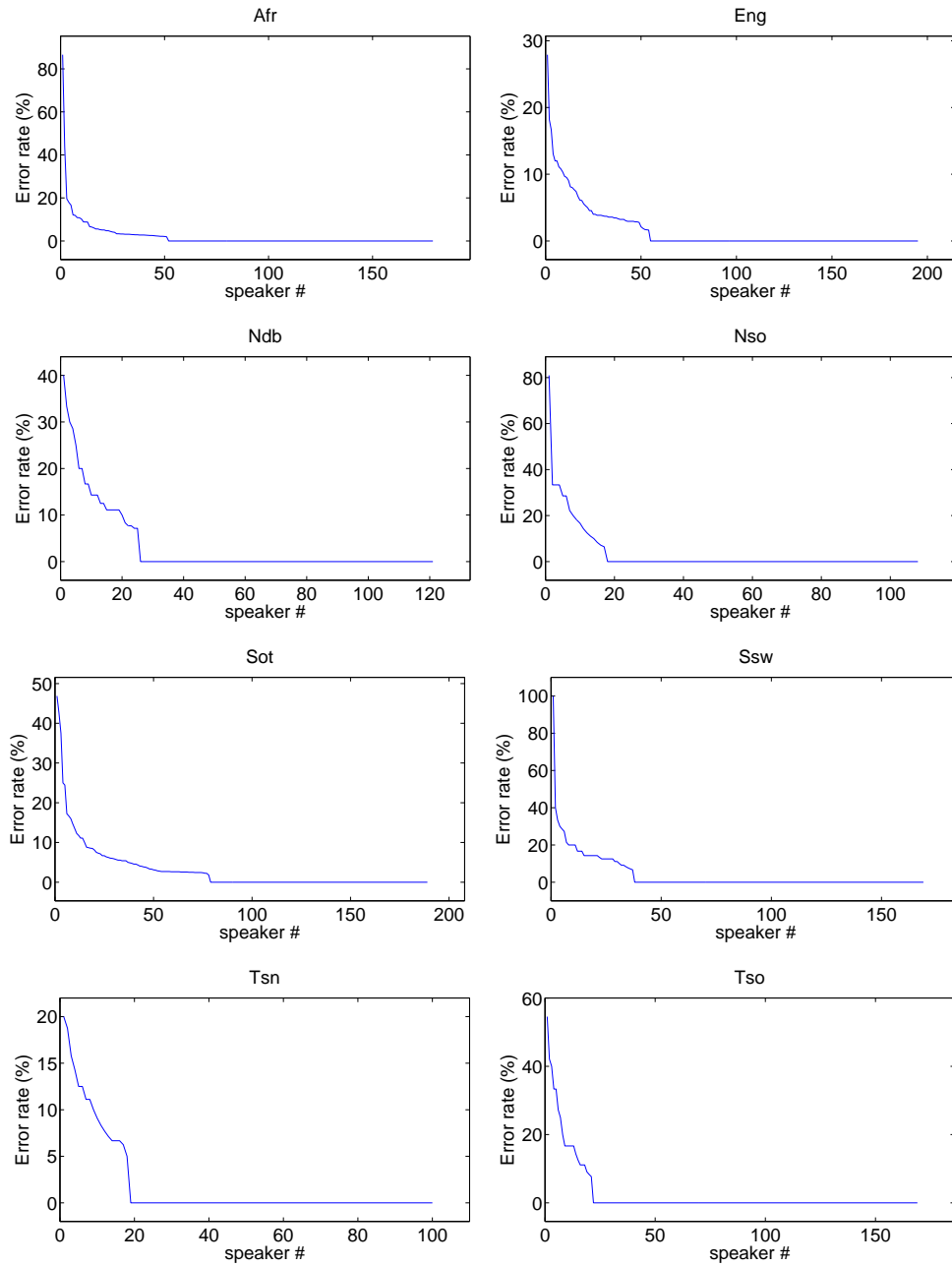


Figure 2.7: Speaker-specific error rates, with speakers arranged from highest to lowest error rates.

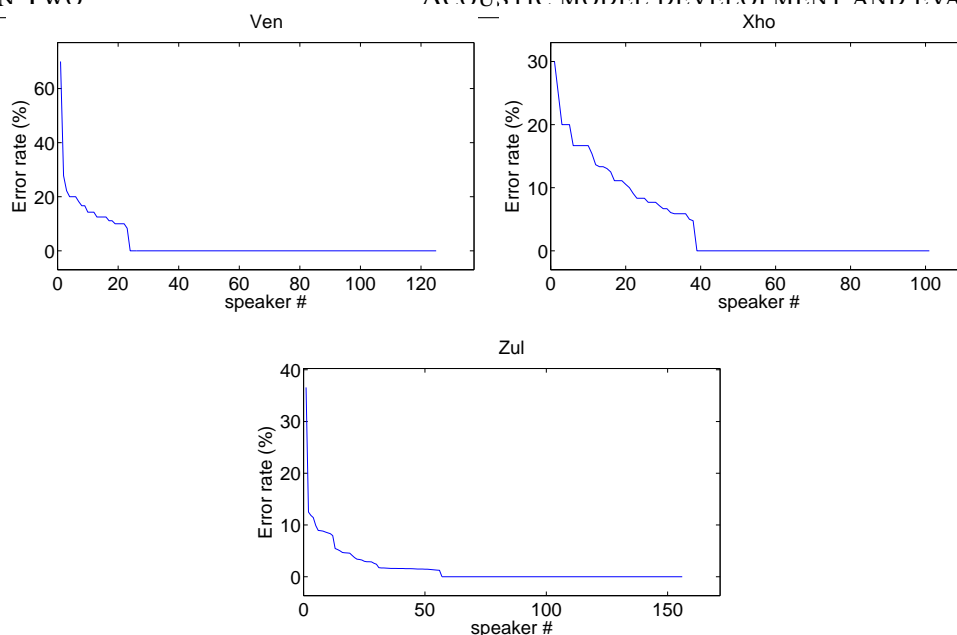


Figure 2.8: Speaker-specific error rates, with speakers arranged from highest to lowest error rates. (*continued*)

raphy of the languages (i.e. abbreviations, acronyms, names and foreign words) is generally much lower than that of standard terms.

- The accuracies achieved do not depend strongly on the amount of training data available for a particular term or the length of the term in phonemes.
- The perplexity of the recognition grammar predictably correlates quite well with the error rate observed.

Our analysis of the speaker-specific performance suggests that the corpus for each language contains a few “goats”, who for some reason or another did not produce recognisable speech. (The term “goats” are often used in speech recognition nomenclature to refer to speakers who produce speech that is particularly hard to recognise.) These speakers should probably be excluded from the corpus, since their problems are typically interface issues rather than recognition failures. Excluding the three worst speakers in each corpus has a significant effect on the overall error rates, as we show in Table 2.4.

Table 2.4: Overall error rates achieved when all speakers are retained, and when the three speakers with the highest error rates are removed.

	Eng	Ndb	Afr	Zul	Sot	Nso	Ssw	Tsn	Tso	Ven	Xho
All speakers	1.76	3.39	3.04	1.52	3.57	4.62	3.85	2.40	3.55	4.31	4.33
3 removed	1.45	2.59	1.56	1.29	2.62	2.46	3.23	1.73	2.39	3.16	3.90

In general, though, the fraction of erroneous utterances in our corpus seems to be quite low, and we suspect that the quality of the trained acoustic models will not be affected much by our corpus-cleaning exercise (or even a manual verification of all utterances). It is expected that additional improvement from an accuracy perspective can be obtained by integrating the outputs from a number of ongoing research projects related to the systematic improvement of pronunciation models, some of which are also being conducted under the auspices of the Lwazi project. These projects include post-graduate research related to the effect of speaker language on name pronunciation, developing acoustic models for code-switched speech, modelling of multilingual variants, text-based language identification of proper names, and the analysis of grapheme-based systems.

In conclusion, we have shown that practically useable speech recognition for our target languages and vocabulary sizes can be achieved with the tools currently at our disposal. The recognisers developed were shown to be vocabulary independent, meaning that – for a similar recording environment – a new application can be developed with minimal effort. For all eleven of South Africa’s official languages, accuracies of higher than 95% were achieved on data that included disfluencies, repetitions and reading errors. If the three worst test speakers were excluded from the test set, an average accuracy across languages of 97.6% is obtained. This accuracy is more than sufficient for deploying speech recognition applications.

SECTION THREE

DEVELOPMENT OF ASR TOOLS

3.1 INTRODUCTION

In this section, we summarise recent improvements to the main acoustic modelling tools and list the current capabilities of the tool set currently referred to as *asr_template*.

3.2 RECENT IMPROVEMENTS

The following improvements to the main set of ASR tools were performed during the past two years:

1. **Better experimental framework:**

- A new set of scripts was created to train and evaluate acoustic models for a given language in an automated fashion (once a group of variables required for task specification have been set).
- All text preprocessing variables (such as dictionaries, punctuation files, phone mappings) now specified in a standard way, and in a single location per experiment.
- The process to split each dataset into a specified number of folds now automated.
- Speaker gender (male and female) now balanced across folds.
- Any missing transcriptions or audio files are automatically removed from processing lists.

2. **Improvements to enable processing of larger corpora:**

- Acoustic model training now parallelised (with the number of processors customisable) and parallelisation of decoding improved. Recombination of results post decoding integrated with main scripts. All code verified to ensure results identical to single processor training/decodes.
- Initial means and variances are calculated on a subset of randomly selected files (currently set at 5000). This provides an immediate speed-up with no loss in performance.
- Search beam width speed and accuracy trade-off analysis for phone recognition resulted in a beam width choice of 70.
- Training all 4 folds of an approximately 80-hour corpus can now be performed in approximately 3 hours.

3. General improvements and bug fixes:

Detail of bug fixes are captured in subversion logs. These include:

- Better handling of diacritics in HTK-generated files.
- Better handling of symbolic links.
- Better handling and logging of errors.
- Trace variables now configurable.
- HTK confusion matrices can now be extracted for large recognition tasks (“MLF files”) as well. (Patch to HTK code, used by asr template.)
- Additional comments, warnings and debugging information added in general.

All changes to the core code/scripts are available under *hlt/asr/asr_template*. An example of the new experimental framework can be accessed in subversion under *hlt/asr_2011/*.

3.3 CURRENT CAPABILITIES

This section contains a brief summary of the current capabilities of *asr_template*. This is a framework of mainly Bash and Perl scripts that makes the process of training ASR acoustic models using HTK simple and efficient, while also providing an environment within which experts can easily add and test new features or templates.

The template has 4 main components:

1. A *configuration file*, *Vars.sh*, where all parameters are set. Currently, almost all parameters are set here (the next release should ensure that all parameters are defined here)

The other 3 components all use the variables set in the configuration file. Furthermore, they should be used in the following hierarchy (the distinction is in practise not as clear as it should ideally be; and is in the process of being improved):

```

--- user callable bash scripts
    --- bash templates
        --- core perl, bash, c scripts

```

where user callable scripts only call templates and templates perform a specific function by among others calling the core scripts.

2. A series of *experiment templates*, for example:

- `mono_train.sh` is a Bash template for training monophone acoustic models
- `tri_train.sh` is a Bash template for training triphone acoustic models
- `semit.sh` is a Bash template for estimating semitied transforms

3. A series of *user callable scripts* in one, non-expert friendly script, for example:

- `mono_train.sh` is a Bash template for training monophone acoustic models
- `TRAIN.sh` allows the user to call different individual or a series of templates to train different parts of an ASR system. It also provides an "all_phases" option, where all components necessary for a functioning standard off-line ASR system are estimated
- `CHECK.sh`, which performs various checks before the actual training commences. This script tries to identify common pitfalls, such as:
 - broken audio files
 - phonemes which have too few training examples
 - words without pronunciations
 - empty transcriptions (may happen after preprocessing)
 - characters in the dictionary (such as digits in the wrong place) which are known to break the HTK tools.

4. A series of core Perl, Bash and C scripts which perform the "low-level" functions, for example:

- `htksortdict.c` sorts a dictionary by string length
- `create_wordlist.pl` creates a unique wordlist from a list of transcriptions
- `create_monophone_list.pl` create a list of monophones from the dictionary

Other advantages over calling HTK directly include:

- These tools exploit HTK's parallelisation capabilities; the user can specify any number of parallel processes; `asr_template` takes care of creating parallel lists, waiting until all sub-processes are done before using the accumulators to estimate the new model when parallelising `HERest`, etc.

- A directory structure and all necessary configuration files and lists are automatically created.
- HMM model numbers are kept track of automatically.

SECTION FOUR

CONCLUSION

This report summarised the progress made with regard to ASR tool development, as well as acoustic model development and evaluation. Acoustic models were generated for all eleven South African languages; these models are able to achieve the target of 95% accuracy on a fairly random 200-term vocabulary. This means that performance is now at a level where the use of ASR in live applications becomes feasible.

APPENDIX A

DELIVERABLES: README

This section contains pointers to all the main deliverables associated with this report: the dictionaries, processed transcriptions and acoustic models.

 DICTIONARIES (copies at ./dictionaries)

language	md5sum	#words	#uniq	ip:path
afrikaans	80402fc5ede581cfc1291cac32a7c036	4302	3835	echelon:/mnt/scratch/drop/lwazi/data/afrikaans/all.dict
english	95a418ab457d4610171e316974b8a60d	7775	7619	echelon:/mnt/scratch/drop/lwazi/data/english/all.dict
isindebele	a064803d58d11c14bad26357fd125616	20051	17666	echelon:/mnt/scratch/drop/lwazi/data/isindebele/all.dict
isizulu	0dc255ece9c4a4dea7315bfebad958a4	3178	2793	echelon:/mnt/scratch/drop/lwazi/data/isizulu/all.dict
sesotho	96df9b0a6e4650419160857d09b313d1	3612	2660	echelon:/mnt/scratch/drop/lwazi/data/sesotho/all.dict
sepedi	a36480d9f15c41f8e9722d673f9684cc	14314	12581	echelon:/mnt/scratch/drop/lwazi/data/sesotho/all.dict
setswana	ff5a3d4cdb86cc600acbe68306a7f233	7773	6212	echelon:/mnt/scratch/drop/lwazi/data/setswana/all.dict
siswati	2f6dc9397ba7a91a29ef93f3ce451371	17540	15332	echelon:/mnt/scratch/drop/lwazi/data/siswati/all.dict
tshivenda	d35284974cee248637d641e15e9a9b7b	11014	9478	echelon:/mnt/scratch/drop/lwazi/data/tshivenda/all.dict
xitsonga	6d63d0dae2b44d358e0021796df97f5c	10939	8897	echelon:/mnt/scratch/drop/lwazi/data/xitsonga/all.dict
isixhosa	9d38bad61a326577291cd665804bff24	31684	30046	echelon:/mnt/scratch/drop/lwazi/data/isixhosa/all.dict

```
-----
PROCESSED TRANSCRIPTIONS (copies at ./processed_transcriptions)
-----
```

```
md5sum: find processed_transcriptions/ -iname "*.txt" | xargs md5sum | sort -u | md5sum
```

language	#files	md5sum	ip:path
afrikaans	99620	2ebe44e5d84f3d2df19af0635ab3a42a	echelon:/mnt/scratch/drop/lwazi/data/afrikaans/processed_transcriptions
english	111374	1a39215f229b1b7c213fd09324742db4	echelon:/mnt/scratch/drop/lwazi/data/english/processed_transcriptions
isindebele	90297	5c3765cec2dc29c5e9840e6d31954528	echelon:/mnt/scratch/drop/lwazi/data/isindebele/processed_transcriptions
isizulu	82552	d478c87ff0bb5cfff5f58f629250a28d	echelon:/mnt/scratch/drop/lwazi/data/isizulu/processed_transcriptions
sesotho	106550	de610e89d892f3a9af112174e087fa48	echelon:/mnt/scratch/drop/lwazi/data/sesotho/processed_transcriptions
sepedi	66959	3e46c1616c675a4daf917f19705ed913	echelon:/mnt/scratch/drop/lwazi/data/sesotho/processed_transcriptions
setswana	65725	0c30e831350dadd37e95ae56b4d8f7c8	echelon:/mnt/scratch/drop/lwazi/data/setswana/processed_transcriptions
siswati	123667	c5e333c3f8edf0b2508f6d202b4d8dd3	echelon:/mnt/scratch/drop/lwazi/data/siswati/processed_transcriptions
tshivenda	119614	d0c3f5830fa9e72d32d63c08834f9fb3	echelon:/mnt/scratch/drop/lwazi/data/tshivenda/processed_transcriptions
xitsonga	121456	64f580664765733961f965af02d5d371	echelon:/mnt/scratch/drop/lwazi/data/xitsonga/processed_transcriptions
isixhosa	76244	534a24deaf6dc2838795d6933c13bb21	echelon:/mnt/scratch/drop/lwazi/data/isixhosa/processed_transcriptions

ACOUSTIC MODELS

afrikaans

f21e7b2a3618ac46401cb1265d614432
d4d1c5d1ed7808d601d1d80c75fa0e56
beb4dedbe2ec567674a8b82e46414ae3
45781c09a84b6a00cbcdc406371e53ee
f7ae92ad7d223270ed76c22662df9cda
fb8c5fc1994559bac8a2e092a90422e9
97a8d58d3588834a7e49da29592e6e48
048d76707e3b9b7d46e5a816b3238b02
88bd071643646c5256a495acde2fac1e
e44d33c3b7089c1a93a7b731e7771031
9f4998b4dbac327fa074a45ff2f926a2
2e4a22de8573223f37c5dc27d8e60274
d1c00d2b3d127bdded76b7321a728631
0c37a0a55441c5276d4172414031a777
ef7673769fcbfd51d8fb9f7a30ffa0a4
00cf44e81af0312f39651ea31246d0a4

echelon:/mnt/scratch/drop/lwazi/data/afrikaans/1/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/1/macros
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/1/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/1/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/2/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/2/macros
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/2/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/2/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/3/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/3/macros
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/3/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/3/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/4/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/4/macros
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/4/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/afrikaans/4/regtree_2.base

english

4438a40c75dfcf85f682a2b4825edfa8
1f91f0365c29c02e06fc6e23557256bb
64ab7954bedced0528f2939e560a6974
b56a6fc05fd1847cac72583fa79d5075
3b4c10ca23e3b1e5b47669bbed30da84
7ccfb51b157a4f6a500a84a5fa684e87
e392620ccb08610a74b559f6e71ab83d
f5fd6b06468e3324e4884946de09505f

echelon:/mnt/scratch/drop/lwazi/data/english/1/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/english/1/macros
echelon:/mnt/scratch/drop/lwazi/data/english/1/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/english/1/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/english/2/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/english/2/macros
echelon:/mnt/scratch/drop/lwazi/data/english/2/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/english/2/regtree_2.base

751f815e6d8e3691db51a335dafb7383
 7744a69745dca0e5a80732dfc0343ef1
 528af36a1af7106957213ebd0448d40f
 1f55d436901bbe5795bdf0254e1046c4
 a5e5043e2cb0bc90ad0105fa0cd499dd
 e8a45b2affe234e1fd929814d42e44a6
 2e24aaad974a4db9f1c891809b509745
 0586480a8116f24bf6ae4b6b1cdb1fec

isindebele

cf5eddd099a573e89cc7c178025822fb
 388ff6b7b21ea1a99f6be8615ae8923f
 0625d4a54fda3a53705f491ecbcc5488
 9ace4db3d089d040995e63c3c3b2f6a0
 5d5fa27c7c04a5b78992e0c1b2a88199
 00bf665c86ce9222cd6ff02919af1171
 d0aa06b88098e70367dbead3fd0b6732
 fa6cb60ee2adfccdf797a1a5deef2ca8
 beaba1307eec74752a0db279d9e03500
 721f07921e97548634b3a783e2c3d172
 cd0e26f0d7b5a211414df8f50667c27f
 e29a52449ec477c974bbf60bef15484f
 9d947c15f7c1464f7fc8e6ec98aa1704
 15c46f1c88e8d6dc8b424f079ba4ef9b
 8ea2b1d60c3c0f5ec1f5dd15eb8c643e
 a837b4525be44f5c0c5f52ab4b26d168

isixhosa

7ccf1bea0772cd5343dalb38c6d883fa
 e8a7939cbf58e8f766509fb10aae90c5

echelon:/mnt/scratch/drop/lwazi/data/english/3/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/english/3/macros
 echelon:/mnt/scratch/drop/lwazi/data/english/3/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/english/3/regtree_2.base
 echelon:/mnt/scratch/drop/lwazi/data/english/4/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/english/4/macros
 echelon:/mnt/scratch/drop/lwazi/data/english/4/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/english/4/regtree_2.base

echelon:/mnt/scratch/drop/lwazi/data/isindebele/1/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/1/macros
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/1/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/1/regtree_2.base
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/2/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/2/macros
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/2/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/2/regtree_2.base
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/3/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/3/macros
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/3/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/3/regtree_2.base
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/4/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/4/macros
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/4/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/isindebele/4/regtree_2.base

echelon:/mnt/scratch/drop/lwazi/data/isixhosa/1/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/isixhosa/1/macros

f9ed0b1391847d71d15f4fe600c1750a
dd0045802dc8ab2a5510c5b560cec798
cb5729fd5f3753b149e46e3126a681cb
bdae37e225df160ef268c47a2986f7d2
7d457156e506ca65cabdb7e81876049b
5fa3061067ae7e8870eed767eda8f3a1
5ac5e30d2a582eb6b70a59d43eb523ce
bae429532d8acdae90530b4133e018f6
c062901cbd16302494df66535c5c04eb
974e830cd8b2af2b096024a07bc720d7
4cc461d9992624f65824701d30fce5f6
ed5daab04464f78c3faf971f2023ac2a
f50998651f12b28e35cd4adb542ab38d
81397f07063fae201776388f1180425e

isizulu

8ebdd27b1737f74803a12dd860132d85
d3e7a8569f54eba3fae1e6dfd591417
5f9afffaa5bf9a000f185c5aad8ad940
67c34ad687e14c60fc9d68a7b7ec9e38
3606a1f54528654966a4bef5c88d8969
6e04ed065ee9e53e82179bff2ba520bb
e58c93af1ee0e6ccbc0f3de35cde7bca
9832cdd5800b950489aafb4b0e7b4c35
dd10c8654a15bc4c3bbaabc735ed203c
d18e8173a6867c817f872383231eb5e1
8bcc3d01a97d7342bf2fd062b0de8c4b
ba565513711324a219cea8481930bfd8
06a10c112b70340139650e34fe1000d3
bb2804a711e7ff634a765206f839266b

echelon:/mnt/scratch/drop/lwazi/data/isixhosa/1/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/1/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/2/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/2/macros
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/2/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/2/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/3/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/3/macros
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/3/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/3/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/4/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/4/macros
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/4/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/isixhosa/4/regtree_2.base

echelon:/mnt/scratch/drop/lwazi/data/isizulu/1/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/isizulu/1/macros
echelon:/mnt/scratch/drop/lwazi/data/isizulu/1/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/isizulu/1/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/isizulu/2/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/isizulu/2/macros
echelon:/mnt/scratch/drop/lwazi/data/isizulu/2/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/isizulu/2/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/isizulu/3/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/isizulu/3/macros
echelon:/mnt/scratch/drop/lwazi/data/isizulu/3/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/isizulu/3/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/isizulu/4/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/isizulu/4/macros

feb3cb2a6cbd36221a4e9085222eff99
621f1400a4ab548ca3cf8843e93f3ef1

sepedi

0ff2f35cda7b9d0721e8771f40d6c5dd
c6856a18d244f8e218ff4838cbf3e481
8d2094632f8769e7b6d7897c8213e476
2728afb8f1193bf9c263ecb904f8fa9c
d4228093a2b1df4081a5e34dbf8dd7a9
03ca90cad4cc51628fd47461db0f6c73
533880b4a42b2d1e84067a2eb4ff7e3c
739956fea98b4697c959895e0a590ae6
07347992cb9b6c5f2ad6c603d1957fc4
945312e144ccacf9210dc2552ff314c
4c873a285cb8c7e29035404535a78362
39ca2c191759b7ac94168060747ce10c
9b5bc509f3290bd4976a5c9fec63b8e9
f548ce5fc566f2ce97892bc6f9612c6c
d98a289c7e62b637043bc350c5822c70
a0d82dfbfb28f3d2bb962b4eaa200c91

sesotho

b79625aa834b707d717fccccd795a44f
b1571417277abedf239447a52492a1e1
fad597f3d8e835cfd2b8ald2692d4204
0d7ccb996e36578ee41a11402086aca2
6c1dd7a38f198e6a9121db2580aa73b1
e2539d47de820095d0831c9368031e7d
ba5ddf3a5016e8eab2d056e1cd35d97c
7b258f77b7b7b4bc3363cb39e9ed7f97

echelon:/mnt/scratch/drop/lwazi/data/isizulu/4/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/isizulu/4/regtree_2.base

echelon:/mnt/scratch/drop/lwazi/data/sepedi/1/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/sepedi/1/macros
echelon:/mnt/scratch/drop/lwazi/data/sepedi/1/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/sepedi/1/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/sepedi/2/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/sepedi/2/macros
echelon:/mnt/scratch/drop/lwazi/data/sepedi/2/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/sepedi/2/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/sepedi/3/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/sepedi/3/macros
echelon:/mnt/scratch/drop/lwazi/data/sepedi/3/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/sepedi/3/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/sepedi/4/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/sepedi/4/macros
echelon:/mnt/scratch/drop/lwazi/data/sepedi/4/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/sepedi/4/regtree_2.base

echelon:/mnt/scratch/drop/lwazi/data/sesotho/1/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/sesotho/1/macros
echelon:/mnt/scratch/drop/lwazi/data/sesotho/1/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/sesotho/1/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/sesotho/2/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/sesotho/2/macros
echelon:/mnt/scratch/drop/lwazi/data/sesotho/2/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/sesotho/2/regtree_2.base

a5d49dbc9d2f293ddcbafcl1a86fd24d4
 697ea7d58234e58944d9964c5b62a09c
 1de0c5dd425d9b8f95b53c0b86b897a0
 b8fdcdca960812c4e6b85e5dc9470475d
 2ca08e25000d23c5b3b40851b974dcf1
 87c81541ddcf55def817aad6d0ef0773
 2683ab616e58f0d4d71702d41bc11ce0
 dde65d9c3b882d86fca32fac4d9f4c64

setswana

84ca7e41077c64159b294b9a663eb932
 60b31602738a9a0cbb6343e10152343e
 b92f520b93256816fa8ffa66b24ddeb4
 e8c0fe10d9404142bd32bbdfdbf869a7
 0cbf4845ee458ddd03299e2ecab86af2
 c8fde6e7ef50edadd71accddde6a2b63
 dc06bfc7f1bf0fce2e033051b50f8a2e
 9524e85d00fd88c574acadeb558b4bf3
 a0569e39edf2a2c6d17285fe52ee742e
 2cddd82da357231e7a65d3625aafd030
 514a187eb91b11fba344f56f7b9a6ff1
 88eb3b71434edf5a94418f9dd52afe36
 3b474df0505508104d47141c07ca9d48
 92f6e2b60680f5e13a587ecd7a5ab01b
 dl60f589404a53ad6ba28cd74b2f2dfb
 dc9ff4caafe0e1eb4250e3e94f82b99d

siswati

a6dd317ecb067a3f3ac09d078b3203a9
 60b065e534c3b645bb9be0f786ef5826

echelon:/mnt/scratch/drop/lwazi/data/sesotho/3/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/sesotho/3/macros
 echelon:/mnt/scratch/drop/lwazi/data/sesotho/3/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/sesotho/3/regtree_2.base
 echelon:/mnt/scratch/drop/lwazi/data/sesotho/4/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/sesotho/4/macros
 echelon:/mnt/scratch/drop/lwazi/data/sesotho/4/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/sesotho/4/regtree_2.base

echelon:/mnt/scratch/drop/lwazi/data/setswana/1/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/setswana/1/macros
 echelon:/mnt/scratch/drop/lwazi/data/setswana/1/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/setswana/1/regtree_2.base
 echelon:/mnt/scratch/drop/lwazi/data/setswana/2/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/setswana/2/macros
 echelon:/mnt/scratch/drop/lwazi/data/setswana/2/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/setswana/2/regtree_2.base
 echelon:/mnt/scratch/drop/lwazi/data/setswana/3/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/setswana/3/macros
 echelon:/mnt/scratch/drop/lwazi/data/setswana/3/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/setswana/3/regtree_2.base
 echelon:/mnt/scratch/drop/lwazi/data/setswana/4/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/setswana/4/macros
 echelon:/mnt/scratch/drop/lwazi/data/setswana/4/SEMITIED
 echelon:/mnt/scratch/drop/lwazi/data/setswana/4/regtree_2.base

echelon:/mnt/scratch/drop/lwazi/data/siswati/1/hmmDefs.mmf
 echelon:/mnt/scratch/drop/lwazi/data/siswati/1/macros

935b06b759e90580b7f2c7230dadae70
8e25218a5b37e8064a24d9c2089f90d9
29de3d7203f9b26ff4b36701837c8b2b
074231a58alac2bd7970clefddf6c526
33d03b2189241a2f99780730486785f3
f7b8825b1f3eeb7f0c977a4b77360467
531a04d6deeccf0fe69b07cb69a3cdd4
88e66d18e643a99407d39246ab2140ef
9021be285882be9dc7428f85652f2d39
a2f77774a28271c7571d147c3bdf231
bcee89080e654a493034f129eefcddb8
ebb6474822305efb5ee7513a75652310
b669d0d11f4855a7f6b3cd8bcd864769
fefe996b0a599aa2f41a609431741c22

tshivenda

dd95868d1f6d9a1b814a39309caa4703
2bf68ec286bb5693a209c6ed8588534d
95bac426aba550e7c3b812736f909f66
c824c461f93b1fffb5bbb01d11a853eda
4dfabe8d6b64347768eb54535bc575ac
996a71f5faa797a63d7a50330657f893
a4411fe40db0e8c18332aa8edae27e88
837de58d6faac9f1e17be42373437884
b81cffb2b735bfbbfd13e57d0484aa14
cc1614c8488309e4709f2036cbbcdcea
15cf8167eb7a780229d468e67f671bb7
dbd0898a20f11fa7ba440fc3f9345757
328f46ec2f3bc981b3e6e2cea5f531d2
c2fadfad1f76f49e9030812elaace8e1

echelon:/mnt/scratch/drop/lwazi/data/siswati/1/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/siswati/1/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/siswati/2/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/siswati/2/macros
echelon:/mnt/scratch/drop/lwazi/data/siswati/2/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/siswati/2/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/siswati/3/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/siswati/3/macros
echelon:/mnt/scratch/drop/lwazi/data/siswati/3/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/siswati/3/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/siswati/4/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/siswati/4/macros
echelon:/mnt/scratch/drop/lwazi/data/siswati/4/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/siswati/4/regtree_2.base

echelon:/mnt/scratch/drop/lwazi/data/tshivenda/1/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/1/macros
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/1/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/1/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/2/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/2/macros
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/2/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/2/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/3/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/3/macros
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/3/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/3/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/4/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/4/macros

38ef44590590dadcd4809c71008cd61e
ede0a8b0683020f2898fbf2f67432e41

xitsonga

879d452581ed6433e48a340d11e5c4df
3fe7b1c4752a234e5efd49b65ca12ff2
b438bb3988e502c09695ca96eb344989
afc0aac1d9e1db1fdc5bb6a2be19d111
a4dfe2f07197a4df65c4c850b265a31f
27c8129a263b3fab32d780f7617b658a
a3880c7c3f671b46134a30e6123f48bc
501ce17f2ddaed1d8afda6ab16fbcff
04373adf589909bcfff0609e6fd2cbd0
44776daed03fad25096e4b28ad582e07
5d60f2a3acade4ca0dd22305dac01f6f
d2eb182d9963f2cd60561b06b6e88d03
3a1d268a2ef599ceaa6af55860a211aa
7a2f1d53c977c3bbcffd658f30d505bf
75a8221694fd07f213404ec5132db707
5fbd03705a8a7b8ddcee4f680d9e295d

echelon:/mnt/scratch/drop/lwazi/data/tshivenda/4/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/tshivenda/4/regtree_2.base

echelon:/mnt/scratch/drop/lwazi/data/xitsonga/1/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/1/macros
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/1/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/1/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/2/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/2/macros
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/2/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/2/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/3/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/3/macros
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/3/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/3/regtree_2.base
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/4/hmmDefs.mmf
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/4/macros
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/4/SEMITIED
echelon:/mnt/scratch/drop/lwazi/data/xitsonga/4/regtree_2.base

REFERENCES

- [1] Marelle H. Davel, Charl van Heerden, and Etienne Barnard, “Medium-Vocabulary Speech Recognition for Under-Resourced Languages,” in *Spoken Languages Technologies for Under-Resourced Languages*, Cape Town, South Africa, May 2012, pp. 146–151.
- [2] N. Patel, D. Chittamuru, A. Jain, P. Dave, and T.S. Parikh, “Avaaj Otalo – A Field Study of an Interactive Voice Forum for Small Farmers in Rural India,” in *Proceedings of the 28th international conference on Human factors in computing systems (CHI 2010)*, Atlanta, GA, USA, April 2010, ACM, pp. 733–742.
- [3] E. Barnard, M.H. Davel, and G.B. van Huyssteen, “Speech Technology for Information Access: a South African case study,” in *Proceedings of the AAAI Spring Symposium on Artificial Intelligence for Development (AI-D)*, Palo Alto, CA, March 2010, pp. 8–13.
- [4] J. Sherwani, S. Palijo, S. Mirza, T. Ahmed, N. Ali, and R. Rosenfeld, “Speech vs. Touch-tone: Telephony Interfaces for Information Access by Low Literate Users,” in *Proc. IEEE Int. Conf. on ICTD*, Doha, Qatar, April 2009, pp. 447–457.
- [5] C. van Heerden, E. Barnard, and M. Davel, “Basic speech recognition for spoken dialogues,” in *Proc. Interspeech*, Brighton, UK, September 2009, pp. 3003–3006.
- [6] C. van Heerden, N. Kleynhans, E. Barnard, and M. Davel, “Pooling ASR Data for Closely Related Languages,” in *Spoken Languages Technologies for Under-Resourced Languages*, Penang, Malaysia, May 2010, pp. 17–23.
- [7] T. Hughes, K. Nakajima, L. Ha, P. Moreno, and M. LeBeau, “Building transcribed speech corpora quickly and cheaply for many languages,” in *Proc. Interspeech*, Makuhari, Japan, September 2010, pp. 1914–1917.
- [8] N.J. De Vries, J. Badenhorst, M.H. Davel, E. Barnard, and A. De Waal, “Woefzela—an open-source platform for asr data collection in the developing world,” in *Proc. Interspeech*, 2011.
- [9] E. Barnard, M. Plauché, and M. Davel, “The Utility of Spoken dialog systems,” in *Proceedings of the 2008 IEEE Workshop on Spoken Language Technology (SLT)*, Goa, India, December 2008, IEEE, pp. 13–16.

- [10] E. Barnard, J. Schalkwyk, C. van Heerden, and P.J. Moreno, “Voice Search for Development,” in *Proc. Interspeech*, Makuhari, Japan, September 2010, pp. 282–285.
- [11] M.H. Davel, C. van Heerden, N. Kleynhans, and E. Barnard, “Efficient Harvesting of Internet Audio for Resource-Scarce ASR,” in *Proc. Interspeech*, Florence, Italy, August 2011, pp. 3153–3156.
- [12] Marelie Davel, Charl van Heerden, and Etienne Barnard, “Validating Smartphone-Collected Speech Corpora,” in *Spoken Languages Technologies for Under-Resourced Languages*, Cape Town, South Africa, May 2012, pp. 68–75.

Annexure I – Requirements analysis templates



<CLIENT/APPLICATION>

Functional Specification Document

1.0

<Date>

Speech Applications Research Group

Meraka Institute



Title	<Client> Functional Specification Document
Version	1.0
Commentator(s)	Gerhard van Huyssteen
Status	Commented Version
Date	2013-02-15
Distribution	HLT GROUP ONLY
Key Words	Needs Specification document, client business requirement, end-user requirements, application requirements, and functional specification document
Project	Lwazi II: XXX

Version	Date	Status	Author/Commentator
Version of document when opened	Date of document as specified by previous author	Status of document when opened	Name of previous author(s) or commentator(s)
1.0.0	18 February 2010		G. van Huyssteen

1. SUMMARY OF REQUIREMENTS

<Give brief summary of most important requirements that will be addressed by this application.>

2. GENERAL DESCRIPTION OF APPLICATION

<Give an overview (in the form of a narrative) of potential application.>

3. INFORMATION/LOGIC DESIGN

<Description of organisation of information; What is grouped together, in what sequence? Provide mindmaps.>

4. MENU DESIGN

<Present taxonomy of menu.>

5. HIGH-LEVEL CALL-FLOW DESIGN

<Present call-flow design in Visio format or as whiteboard design; could be attached as addendum>

6. SAMPLE DIALOGUE DESIGN

<Give a few examples of dialogues, in order to capture style and register.>

7. PERSONA DESIGN

<Give overview of persona, without going into too much detail.>

- Number of personas
- Language
- Gender
- Age
- Real vs. synthetic
- Generic vs. unique
- Standard vs. custom

8. UNIVERSAL BEHAVIOUR/GLOBAL COMMANDS

<Give list of universal behaviours, such as accessing operator, going back to main menu, etc.>

9. DESIGN POLICIES

<Describe general points of departure related to personalisation, queuing, outbound strategies, existing strategies, non-speech audio (silent strategy), prompt recordings, DTMF, confidence handling, time-out, etc.>

10. DESIGN PRIORITIES

<What is important to the client? Is it branding? Is it making money?>

- Marketing
- Design
- Productivity
- Aesthetics
- Ergonomics

11. GOVERNANCE

<Give some specifications regarding governance of the proposed application.>

- Multilingual system
- Storage

- Change management
- Subcontractors
- Procedures

12. RISK MANAGEMENT

<Provide a list of potential risks, and how these risks could be addressed.>



<CLIENT/APPLICATION>

Needs Definition Document

1.0

<Date>

Speech Applications Research Group

Meraka Institute



Title	<Client> Needs Definition Document
Version	1.0
Commentator(s)	Gerhard van Huyssteen
Status	Commented Version
Date	2013-02-15
Distribution	HLT GROUP ONLY
Key Words	Needs Specification document, client business requirement, end-user requirements, application requirements, and functional specification document
Project	Lwazi II: XXX

Version	Date	Status	Author/Commentator
Version of document when opened	Date of document as specified by previous author	Status of document when opened	Name of previous author(s) or commentator(s)
1.0.0	18 February 2010		G. van Huyssteen

1. THE NEED

<Give a short narrative expressing the need, relating to the end user>

2. THE CLIENT

<Define the client; role and context>

3. ADDITIONAL INFORMATION

<Provide additional information (e.g. existing system)>

- Document should be 1-2 pages



<CLIENT/APPLICATION>

Requirements Specification Document

1.0

<Date>

Speech Applications Research Group

Meraka Institute



Title	<Client> Requirements Specification Document
Version	1.0
Commentator(s)	Gerhard van Huyssteen
Status	Commented Version
Date	2013-02-15
Distribution	HLT GROUP ONLY
Key Words	Needs Specification document, client business requirement, end-user requirements, application requirements, and functional specification document
Project	Lwazi II: XXX

Version	Date	Status	Author/Commentator
Version of document when opened	Date of document as specified by previous author	Status of document when opened	Name of previous author(s) or commentator(s)
1.0.0	18 February 2010		G. van Huyssteen

1. CLIENT BUSINESS REQUIREMENTS

1.1. BUSINESS DRIVERS

< Cost saving, user satisfaction, information connection and improving access >

1.2. BUSINESS GOALS PRIORITISED

<Task completion, user satisfaction (increase number of users by a certain percentage) and brand image. Define how you will measure satisfaction.>

2. END-USER REQUIREMENTS

2.1. USER PROFILE

<(WHO?). Demographics of 1-3 typical users.>

- Age group
- Gender
- Language
- Dialect
- Literacy
- Technology literacy (know how to use something)
- Access to technology (own or shared)
- Usage of technology (contract vs. pay-as-go; sms; internet)
- Level of education
- Employment (employed, unemployed; typical verticals)
- LSM (Living standard measure)
- Geographical location (urban/rural)

2.2. USAGE PROFILE

<(WHERE? WHEN? and WHY?). Describe typical user behaviour.>

2.3. USER MENTAL MODEL

< **WHAT** does the user expect? Example: When a person calls a certain individual, s/he expects to talk to that person; if not then to be transferred to the receptionists.>

2.4. USER SUCCESS CRITERIA

<When will the user be 'satisfied' with the system? Example: If task completion is above 80%; consider factors such as speed, ease of use, entertainment value.>

3. APPLICATION REQUIREMENTS

3.1. EXISTING SYSTEMS AND/OR SOLUTIONS

<Existing systems and/or solutions. For example existing persona, back-end systems, etc.>

3.2. TASK ANALYSIS

<List of transactions from system to user and back. Descriptions of intended tasks.>

3.3. FUNCTIONAL/FEATURE REQUIREMENTS

<Breakdown of tasks in features. Examples: Fall-back, queuing, call centre, missed calls, user-ID, message retrieval, message storage, enter message, edit message, etc.>

3.4. APPLICATION ENVIRONMENT

<Hardware platforms, software platforms, telephony, integrating software, supporting technologies (ASR, TTS & barge-in), etc.>

NOTE: Possible interventions:

- Talk to client (interviews; joint requirements planning (JRP) workshop)
- Talk to user (interviews; joint requirements planning (JRP) workshop; questionnaires)
- Meeting with developers/ technical department of the client (interviews)
- Meeting with the marketing department (interviews; workshop)



<CLIENT/APPLICATION>

Technical Specification Document

1.0

<Date>

Speech Applications Research Group

Meraka Institute



Title	<Client> Technical Specification Document
Version	1.0
Commentator(s)	Gerhard van Huyssteen
Status	Commented Version
Date	2013-02-15
Distribution	HLT GROUP ONLY
Key Words	Needs Specification document, client business requirement, end-user requirements, application requirements, and functional specification document
Project	Lwazi II: XXX

Version	Date	Status	Author/Commentator
Version of document when opened	Date of document as specified by previous author	Status of document when opened	Name of previous author(s) or commentator(s)
1.0.0	18 February 2010		G. van Huyssteen

1. DETAILED CALL-FLOW DESIGN

<Specify where detailed call-flow design is located. Supply additional information.>

2. DIALOGUE STATES

<Could be described separately, or as part of detailed call-flow design.>

3. DEVELOPMENTAL NOTES

<Provide additional notes to developers/designers, such as ASR confidence, etc.>

4. GRAMMARS

<Specify where grammars are located. Supply additional information.>

5. PROMPT SHEET/RECORDING SHEET

<Specify where prompt sheet is located. Provide additional information for recording studio and voice artist.>

NOTE: Use “VUI Toolkit” as standard design instrument.

Annexure J – Speech processing pipeline detailed results

		static extracted features					running features
		HTK	Julius				
		16k Sample rate			8k sample rate		
# cnts	Test term	static cmn	static cmn	real-time cmn	static cmn	running cmn	running cmn
1	american pacifists	100.00	100.00	100.00	100.00	100.00	100.00
1	american zoologists	100.00	100.00	100.00	100.00	100.00	100.00
1	apply binary search	100.00	100.00	100.00	100.00	100.00	100.00
1	british special operations	100.00	100.00	100.00	100.00	100.00	100.00
1	calvin_s parents thought	100.00	100.00	100.00	100.00	100.00	100.00
1	external magnetic field	100.00	100.00	100.00	100.00	100.00	100.00
1	fencers use differ	100.00	100.00	100.00	100.00	100.00	100.00
1	french air force	100.00	100.00	100.00	100.00	100.00	100.00
1	government accountability office	100.00	100.00	100.00	100.00	100.00	100.00
1	java libraries	100.00	100.00	100.00	100.00	100.00	100.00
1	larger joints like	100.00	100.00	100.00	100.00	100.00	100.00
1	narrow gauge lines	100.00	100.00	100.00	100.00	100.00	100.00
1	states national film	100.00	100.00	100.00	100.00	100.00	100.00
1	technology nanded sggsie	100.00	100.00	100.00	100.00	100.00	100.00
1	university towns in the united states	100.00	100.00	100.00	100.00	100.00	100.00
1	world economic forum	100.00	100.00	100.00	100.00	100.00	100.00
2	cardiff arms park	100.00	100.00	100.00	100.00	100.00	100.00
2	eastern	100.00	100.00	100.00	100.00	100.00	100.00
2	heian period japan	50.00	0.00	0.00	50.00	50.00	0.00
2	largest tidal whirlpool	100.00	100.00	100.00	100.00	100.00	100.00
3	capitals in europe	100.00	100.00	100.00	100.00	100.00	100.00
3	his name	100.00	66.67	66.67	66.67	66.67	66.67
3	home	66.67	66.67	66.67	66.67	66.67	100.00
4	advocated tory democracy	100.00	75.00	75.00	75.00	75.00	75.00
4	boinae by common name	75.00	50.00	50.00	50.00	50.00	66.67
4	deaths from nephritis	75.00	75.00	75.00	75.00	75.00	100.00
4	humans appear white	100.00	100.00	100.00	100.00	100.00	100.00
4	msd radix sort	100.00	75.00	25.00	25.00	25.00	25.00
4	wild card	100.00	100.00	100.00	100.00	100.00	100.00
5	brusciotto	100.00	100.00	100.00	100.00	100.00	100.00
5	greater london authority	100.00	100.00	100.00	100.00	100.00	100.00
5	international statistical classification	100.00	100.00	100.00	100.00	100.00	100.00
5	league rbi champions	100.00	100.00	100.00	100.00	100.00	100.00
5	old mancurians	100.00	100.00	100.00	100.00	100.00	100.00
5	sid ahmed taya	100.00	100.00	100.00	100.00	100.00	100.00
5	widely used local	100.00	100.00	100.00	100.00	100.00	100.00
6	accelerators provide	100.00	100.00	100.00	100.00	100.00	100.00
6	agreement recently signed	100.00	100.00	100.00	100.00	100.00	100.00
6	best original screenplay	100.00	100.00	100.00	100.00	100.00	100.00
6	filmmakers applied socialist	100.00	83.33	83.33	83.33	83.33	83.33

		static extracted features					running features
		HTK	Julius				
		16k Sample rate			8k sample rate		
# cnts	Test term	static cmn	static cmn	real-time cmn	static cmn	running cmn	running cmn
6	french president nicolas	100.00	100.00	100.00	100.00	100.00	100.00
6	nokia siemens networks	100.00	100.00	100.00	100.00	100.00	100.00
6	saint louis cemetery	83.33	83.33	83.33	83.33	83.33	100.00
6	spruit	66.67	83.33	83.33	66.67	83.33	125.00
7	black hawk war	100.00	100.00	100.00	100.00	100.00	100.00
7	conic equation given	100.00	100.00	100.00	100.00	100.00	100.00
7	french grand prix	100.00	100.00	100.00	100.00	100.00	100.00
7	northern constellations	100.00	100.00	100.00	100.00	100.00	100.00
8	ad hoc basis	100.00	100.00	100.00	100.00	100.00	100.00
8	fiction writers usually	100.00	100.00	100.00	100.00	100.00	100.00
8	language families	100.00	100.00	100.00	100.00	100.00	100.00
9	demographics of armenia	100.00	100.00	100.00	100.00	100.00	100.00
9	phi beta kappa	77.78	77.78	77.78	88.89	88.89	100.00
9	royal institution christmas	100.00	100.00	100.00	100.00	100.00	100.00
9	sharonsville	100.00	100.00	100.00	100.00	100.00	100.00
9	suburban community newspapers	100.00	100.00	100.00	100.00	100.00	100.00
9	twenty	100.00	100.00	100.00	100.00	100.00	100.00
9	victor emmanuel	100.00	100.00	100.00	100.00	100.00	100.00
9	von siebold_s collection	88.89	88.89	88.89	88.89	88.89	100.00
10	central limit theorem	100.00	100.00	100.00	100.00	100.00	100.00
10	clarence_s	100.00	100.00	100.00	100.00	100.00	100.00
10	local government act	100.00	100.00	100.00	100.00	100.00	100.00
10	minister tony blair	100.00	100.00	100.00	100.00	100.00	100.00
10	thirty years war	100.00	100.00	100.00	100.00	100.00	100.00
10	women philosophers	90.00	90.00	90.00	90.00	90.00	100.00
11	ancient greek poets	100.00	100.00	100.00	100.00	100.00	100.00
11	danish musical instruments	100.00	100.00	100.00	100.00	100.00	100.00
11	density functional theory	100.00	100.00	100.00	100.00	100.00	100.00
11	knowledge representation languages	100.00	100.00	100.00	100.00	100.00	100.00
11	sir francis drake	100.00	100.00	100.00	100.00	100.00	100.00
11	south american countries	100.00	100.00	100.00	100.00	100.00	100.00
11	supermassive black hole	100.00	100.00	100.00	100.00	100.00	100.00
12	british north american	100.00	100.00	100.00	100.00	100.00	100.00
12	corps air station	100.00	100.00	100.00	100.00	100.00	100.00
12	lucky	100.00	100.00	100.00	100.00	100.00	100.00
12	san francisco giants	100.00	100.00	100.00	100.00	100.00	100.00
12	winston cup championship	100.00	100.00	100.00	100.00	100.00	100.00
13	computational complexity theory	100.00	100.00	100.00	100.00	100.00	100.00
13	geum river basin	92.31	92.31	92.31	92.31	92.31	100.00
14	american atheists	92.86	92.86	92.86	78.57	85.71	100.00

		static extracted features					running features
		HTK	Julius				
		16k Sample rate			8k sample rate		
# cnts	Test term	static cmn	static cmn	real-time cmn	static cmn	running cmn	running cmn
14	bafta winners people	100.00	100.00	100.00	100.00	100.00	100.00
14	ethiopian orthodox church	100.00	100.00	100.00	100.00	100.00	100.00
14	states virgin islands	100.00	100.00	100.00	100.00	100.00	100.00
16	edmund beck osb	100.00	93.75	93.75	100.00	93.75	93.75
16	include object oriented analysis	100.00	100.00	100.00	100.00	100.00	100.00
17	bishops of worcester	94.12	94.12	94.12	94.12	94.12	100.00
17	bob dylan organist	100.00	100.00	100.00	100.00	100.00	100.00
17	event took place	100.00	100.00	100.00	100.00	100.00	100.00
18	aldrin added	100.00	100.00	94.44	100.00	100.00	94.44
18	french foreign legion	100.00	100.00	100.00	94.44	100.00	100.00
18	minimalism include musicians	100.00	94.44	94.44	100.00	100.00	94.44
18	norro generally administer	100.00	100.00	100.00	100.00	100.00	100.00
18	yogic guru nandhi	88.89	88.89	88.89	94.44	83.33	100.00
19	and crown thy good with brotherhood	100.00	100.00	100.00	100.00	100.00	100.00
19	artificial neural network	100.00	100.00	100.00	100.00	100.00	100.00
19	central american parliament	100.00	100.00	94.74	94.74	100.00	94.74
19	durban	94.74	94.74	89.47	94.74	84.21	94.44
20	chicago white stockings	95.00	95.00	95.00	95.00	95.00	100.00
20	cities in the netherlands	100.00	95.00	95.00	90.00	90.00	95.00
20	eastern orthodox saints	100.00	100.00	100.00	100.00	100.00	100.00
20	ink flowed past	100.00	100.00	100.00	100.00	100.00	100.00
20	insufficient material available	100.00	100.00	100.00	100.00	100.00	100.00
20	mission specialist educators or	100.00	100.00	100.00	100.00	100.00	100.00
20	owain glynd	95.00	95.00	95.00	95.00	90.00	100.00
20	senior high school	100.00	100.00	100.00	100.00	100.00	100.00
20	that saved a wretch like me	100.00	100.00	100.00	100.00	100.00	100.00
21	eastern roman emperor	100.00	100.00	100.00	100.00	100.00	100.00
21	mexican film directors	100.00	100.00	100.00	100.00	100.00	100.00
22	converts to buddhism	100.00	100.00	100.00	95.45	95.45	100.00
23	african origin view	100.00	100.00	100.00	100.00	100.00	100.00
24	complex vector space	100.00	100.00	100.00	100.00	100.00	100.00
24	real world artists	100.00	100.00	100.00	100.00	100.00	100.00
25	federal information processing	100.00	100.00	100.00	100.00	100.00	100.00
25	signal transduction pathways	96.00	92.00	92.00	96.00	96.00	95.83
25	underutilized crops	96.00	96.00	96.00	96.00	96.00	100.00
26	rational choice theory	96.15	84.62	84.62	100.00	100.00	88.00
27	paj czno county	81.48	92.59	96.15	77.78	81.48	118.01
28	early ninth century	96.43	96.43	96.30	96.43	92.86	99.86
29	king christian	96.55	96.55	96.55	96.55	96.55	100.00
29	roman gods	100.00	96.55	96.55	100.00	100.00	96.55

		static extracted features					running features
		HTK	Julius				
		16k Sample rate			8k sample rate		
# cnts	Test term	static cmn	static cmn	real-time cmn	static cmn	running cmn	running cmn
30	national ice hockey	96.67	96.67	96.67	93.33	93.33	100.00
30	united workers party	100.00	100.00	100.00	100.00	100.00	100.00
33	samuel taylor coleridge	96.97	93.94	93.94	96.97	96.97	96.88
35	ptolemaic court	94.29	91.43	91.43	94.29	91.43	96.97
39	ludwig mies van	92.31	89.74	87.18	97.44	84.62	94.44
40	educational institutions established	100.00	100.00	100.00	100.00	100.00	100.00
41	among others john heath has observed	100.00	100.00	100.00	100.00	100.00	100.00
41	arab league member	100.00	100.00	100.00	100.00	100.00	100.00
41	art	100.00	100.00	97.56	90.24	95.12	97.56
41	defense missile squadron	97.56	97.56	97.56	100.00	100.00	100.00
41	german brands	100.00	100.00	97.56	100.00	100.00	97.56
41	meeting martin says	100.00	100.00	100.00	100.00	100.00	100.00
41	mythological greek archers	100.00	100.00	100.00	100.00	100.00	100.00
42	photographer stanley forman	100.00	100.00	100.00	100.00	100.00	100.00
42	television series endings	97.62	100.00	100.00	100.00	100.00	102.44
43	legal documents	100.00	100.00	97.67	100.00	100.00	97.67
43	water	100.00	100.00	100.00	100.00	100.00	100.00
44	french unitarians	100.00	97.73	97.73	100.00	100.00	97.73
46	freedmpark	100.00	100.00	100.00	100.00	100.00	100.00
48	warner home video	97.92	97.92	97.92	97.92	97.92	100.00
49	columbia law school	95.92	97.96	97.96	97.96	100.00	102.13
49	sci fi novel fahrenheit	97.96	97.96	97.96	95.92	95.92	100.00
50	increase flow resistance	100.00	98.00	98.00	100.00	100.00	98.00
51	west bromwich albion	100.00	100.00	100.00	98.04	100.00	100.00
52	astrological sign called	98.08	98.08	98.08	96.15	96.15	100.00
52	granth hib	100.00	100.00	100.00	98.08	98.08	100.00
52	improving international relations	100.00	100.00	100.00	100.00	100.00	100.00
54	allowed foreign nations	100.00	100.00	98.15	100.00	100.00	98.15
54	gdp growth rate	100.00	96.30	98.15	98.15	98.15	98.15
54	luther king jr	98.15	98.15	98.15	96.30	90.74	100.00
54	major league career	100.00	100.00	98.15	96.30	96.30	98.15
54	megafauna of africa	98.15	100.00	98.15	98.15	96.30	100.00
54	preston north end	94.44	98.15	94.44	98.15	96.30	100.00
54	south africa won	100.00	98.15	100.00	100.00	100.00	100.00
55	ancient athenians	100.00	98.18	96.36	98.18	98.18	96.36
55	article relates primarily	96.36	98.18	94.55	94.55	92.73	98.11
55	automatic send receive	100.00	100.00	100.00	100.00	100.00	100.00
55	beneath grade levels	100.00	100.00	100.00	100.00	100.00	100.00
55	ferry	98.18	98.18	98.18	98.18	98.18	100.00
55	hercule poirot characters	96.36	92.73	90.91	92.73	92.73	94.34

		static extracted features					running features
		HTK	Julius				
		16k Sample rate			8k sample rate		
# cnts	Test term	static cmn	static cmn	real-time cmn	static cmn	running cmn	running cmn
55	murdered roman empresses	98.18	98.18	98.18	98.18	96.36	100.00
55	red river rebellion	98.18	98.18	98.18	98.18	98.18	100.00
56	largest natural gas	98.21	100.00	98.21	98.21	98.21	100.00
56	original music score	100.00	100.00	100.00	100.00	100.00	100.00
56	roman emperor constantine	100.00	98.21	98.21	100.00	98.21	98.21
56	santa fe trail	100.00	96.43	96.43	83.93	87.50	96.43
57	software engineering institute	100.00	100.00	98.25	98.25	98.25	98.25
58	cold fusion researchers	98.28	98.28	98.28	98.28	98.28	100.00
58	compatible instruction set	100.00	100.00	100.00	100.00	100.00	100.00
58	high priests of israel	98.28	96.55	96.55	98.28	98.28	98.25
59	indie friendly london records	100.00	100.00	100.00	100.00	100.00	100.00
60	living fossils	96.67	98.33	98.33	98.33	98.33	101.72
61	climate classification cfa	100.00	100.00	100.00	100.00	100.00	100.00
61	congress	96.72	96.72	96.72	96.72	96.72	100.00
61	national basketball league	100.00	100.00	100.00	98.36	98.36	100.00
62	german atheists	98.39	93.55	91.94	95.16	90.32	93.44
64	gay artists	100.00	100.00	98.44	100.00	98.44	98.44
64	press syndicate continued	100.00	100.00	100.00	100.00	100.00	100.00
64	western constellations	98.44	98.44	98.44	98.44	98.44	100.00
65	coronary heart disease	100.00	100.00	100.00	100.00	100.00	100.00
65	features atoms wide	100.00	100.00	100.00	100.00	100.00	100.00
65	right hand drive	100.00	100.00	100.00	100.00	100.00	100.00
66	british world music groups	100.00	98.48	98.46	100.00	100.00	98.46
66	coast main line	100.00	100.00	100.00	100.00	100.00	100.00
66	japanese personal pronouns	100.00	100.00	100.00	100.00	100.00	100.00
66	london symphony orchestra	100.00	98.48	98.48	98.48	98.48	98.48
66	probability density function	100.00	100.00	100.00	100.00	100.00	100.00
66	scholastic philosophers	98.48	96.97	96.97	98.48	96.97	98.46
66	scottish inventors	98.48	98.48	98.48	98.48	98.48	100.00
67	britain_s moorish revival	98.51	98.51	98.51	98.51	98.51	100.00
67	including international competitors	100.00	100.00	100.00	100.00	100.00	100.00
67	local nomadic tribes	98.51	98.51	98.51	98.51	98.51	100.00
67	polish musical instruments	100.00	100.00	97.01	100.00	98.51	97.01
67	pope innocent	98.51	98.51	98.51	98.51	98.51	100.00
67	walt disney world	100.00	100.00	100.00	100.00	98.51	100.00
68	resources include petroleum	100.00	98.53	100.00	100.00	100.00	100.00
68	simple harmonic motion	100.00	100.00	100.00	100.00	100.00	100.00
68	south wales valleys	98.53	98.53	98.53	98.53	98.53	100.00
74	human anatomy	100.00	98.65	98.65	98.65	98.65	98.65
79	union political leaders	98.73	97.47	97.47	97.47	97.47	98.72

Annexure K – Future infrastructures report

Executive summary

In 2010 the HLT research group (HLT RG) of the CSIR, Meraka Institute was awarded a three-year contract from the Department of Arts and Culture (DAC) for a project entitled *Lwazi II: Increasing the impact of speech technologies in South Africa*.

The Lwazi II project follows on from Lwazi I and involves further research and development work on speech technologies for resource-scarce languages, mainly the official languages of South Africa. The project is divided into two sub-projects: *speech technologies* and *speech applications*.

The speech technology development work comprises the development of text-to-speech (TTS) and automatic speech recognition (ASR) resources and systems. The speech applications work relates to the selection, design, development, piloting, monitoring and evaluation of viable speech-based services. Other deliverables include platform development, i.e. the development of a speech processing pipeline and a content management system; a technology transfer master plan; and a future infrastructures report.

This future infrastructures report provides an overview of alternative infrastructures identified and tested over the course of the Lwazi II project. It also indicates how these may be used in future deployments of interactive voice response (IVR) systems to address issues such as scalability, 24/7/365 uptime of services, and rapid design and prototype testing of IVR-based services.

Table of Contents

Executive summary.....	1
List of figures.....	2
1. Purpose.....	3
2. Background.....	3
3. Recommended infrastructures.....	3
3.1 Voxeo.....	3
3.2 Lwazi platform and service hosting.....	4
3.2.1 Bulk-SMS.....	4
3.2.2 IVR hosting.....	5
4. Conclusions and Way Forward	7

List of figures

Figure 1 - External hosting configuration	6
---	---

1. Purpose

The purpose of this report is to provide an overview on future infrastructures that were identified and utilized during the course of the project to complement the various technologies developed.

2. Background

The HLT research group of the CSIR, Meraka Institute developed a number of telephone-based services for the Lwazi II project. Each of these services required replicable telephony and computing infrastructure. The services relied on the availability of telephone lines, SMS gateways, as well as computer servers for running telephony platform and application software.

The majority of these services were run using internal HLT telephony and computing infrastructure. There were two major issues with respect to this:

- Limited ability to scale the services up for potential large-scale uptake by users.
- No guaranteed service uptime (i.e. 24-hour uptime).

With this in mind, an investigation was undertaken to identify and test out viable infrastructure that could be used to address the above issues.

In addition, an investigation was conducted into an alternative platform software and service provider to determine how it could improve the design, development and testing of service prototypes early on with users.

3. Recommended infrastructures

3.1 Voxeo

Voxeo was investigated by the HLT RG as a tool for use during the design phase of IVR application development. This was explored as an alternative to requiring intensive involvement from both application designers and software developers during the creation of prototypes or quick mock-ups of eventual services.

Voxeo (<http://www.voxeo.com/>) provides a number of *attractive features* for this, including:

- A simple to use graphical designer enabling application designers (non-developers) to create an interactive voice response call-flow.
- Service delivery infrastructure: applications developed using the tool could be provisioned on a local telephone number running on hosted infrastructure in the United States and Europe. In essence, application designers register their application with Voxeo to be provided on a local telephone number. Voxeo allocates a local number (in our case a number in South Africa) for the service.

There were, however, also a number of *limiting factors*:

- Provision is only made for incoming calls – no call-back communication model is available to enable free calls for users.
- All calls are forwarded to the service running outside the country - which has the risk of introducing latency (time-delay) issues.
- Difficulty in integrating the HLT RG's automatic speech recognition (ASR) technologies. Significant development would have had to be done for our speech-recognition server to interface with the Voxeo platform.
- Development of fully-fledged services versus the simple design prototypes would still require intensive involvement from software developers, especially for integration with databases.

Due to the above limitations a decision was made to continue using the existing telephony platform software and the existing design, development and testing process.

3.2 Lwazi platform and service hosting

3.2.1 Bulk-SMS

Many of the Lwazi II pilot services required an SMS feedback channel with users. An example would be for the AfriVet services, where SMSes are used to remind vets and dip tank attendants to call into the V-Plan Vet line and the Dip tank line respectively.

A number of SMS service providers who host the required infrastructure and make billing cost effective and convenient, exist. It was decided to make use of such a service provider for our services - in our case we chose Bulk-SMS.

In essence this provided a number of major *benefits*:

- No hardware was required to enable SMS messaging (e.g. GSM gateways).
- No interaction with mobile operators (MTN, Vodacom) was required, i.e. no lengthy contracts or billing. Instead, one simply required a 30-day notice account.
- Service could also be provided at scale, since providers ensured 24-hour availability and more than sufficient volume for sending out messages on-demand.
- All our services that require SMS interface directly with Bulk-SMS servers via the Internet using the well-established HTTP (hypertext transfer protocol), and simple to use API commands provided by Bulk-SMS.

The above solution was implemented successfully for all our Lwazi II services that required an SMS capability.

3.2.2 IVR hosting

The Lwazi II project has required dedicated telephony hosting for the various pilot applications developed over the years (e.g. CDW services, DBE NSNP services, etc.). This was mainly achieved by hosting hardware and software internally using the HLT RG's own resources.

This, however, introduced a number of *challenges* over time, including:

- Running out of outgoing lines in our building (expensive hardware would need to be purchased to increase this capacity)
- High cost of leasing sufficient outgoing lines
- The inability to guarantee 24-hour uptime (no support staff available).

The majority of these challenges could be addressed by approaching external hosted service providers for any new services developed.

Hosting of IVR services requires both telephony and computing hardware. On the telephony end there is a need for telephone lines to enable incoming and outgoing calls. Computer servers are also required to run the Lwazi II platform and application software.

Two external service providers were identified to meet this need - one capable of providing the required telephone lines, and the other one capable of providing servers to host our software services. Switchtel was selected as our telephony provider, and Snowball as our hosted server provider. The following figure is a high level view of how the service is hosted externally.

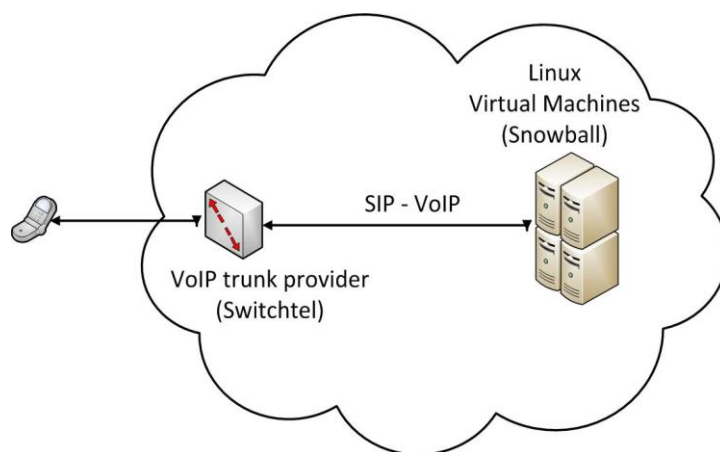


Figure 1 - External hosting configuration

As depicted, phone calls get made via a VoIP (voice over IP) trunk provided by Switchtel. These calls then get forwarded to our software services running on hosted server virtual machines provided by Snowball.

VoIP technology and its related protocols (SIP) is one of the critical enablers for hosted voice service provision. It enables voice calls to be forwarded from the telephone line provider (Switchtel) to our hosted server provider (Snowball) via the Internet. The above external hosting configuration has been successfully employed for hosting new services developed, in particular the AfriVet services.

4. Conclusions and Way Forward

Our investigation into an alternative platform for use during the prototyping phase of IVR service development identified the need for our own platform to be able to cater for this in some form. Future development on the platform should enable the creation of quick and inexpensive prototypes for testing with users.

The abovementioned SMS hosting solution has been extensively applied to the Lwazi II services and provides solid support for any additional services to be rolled out.

Lastly, the IVR hosting configuration has so far proven to be reliable. The solution provides an integral base for scaling our services up for larger scale usage and enhancing the existing software platform to cater for this.