

Etienne Barnard and Marelie Davel

CSIR, Pretoria, South Africa
EBarnard@csir.co.za; MDavel@csir.co.za

LLSTI isiZulu TTS

Evaluation Report

Keywords: *TTS Evaluation, intelligibility, questionnaire creation*

Contents

1	INTRODUCTION	2
2	EVALUATION	2
2.1	GOAL.....	2
2.2	METHODS.....	2
2.3	RESULTS.....	3
2.3.1	E0 Evaluation.....	3
2.3.2	E1 Evaluation.....	4
2.3.3	E2 Evaluation.....	4
3	CONCLUSION	5
	APPENDIX A: TELEPHONY EVALUATION DIALOGUE	6
	isiZULU VERSION OF IVR DIALOGUE.....	6
	ENGLISH TRANSLATION OF IVR DIALOGUE.....	8
	APPENDIX B: TELEPHONY EVALUATION QUESTIONNAIRE	10
	isiZULU VERSION OF EVALUATION QUESTIONNAIRE.....	10
	ENGLISH TRANSLATION OF EVALUATION QUESTIONNAIRE.....	12
	APPENDIX C: WEB-BASED EVALUATION	13
	APPENDIX D: DETAILED RESULTS	15
	E0 EVALUATION.....	16
	E1 EVALUATION.....	17
	E2 EVALUATION.....	17

1 Introduction

As part of the Local Languages Speech Technologies Initiative (LLSTI), the CSIR South Africa has developed an isiZulu text-to-speech (TTS) system, based on the open-source toolkit Festival. An initial version of the system was completed in July 2004, and a pilot evaluation was designed in order to determine its usability.

This report describes the goals of the pilot evaluation, the methods employed and results obtained. Further details, including the full set of results, are included in the appendices.

2 Evaluation

2.1 Goal

A pilot evaluation was designed in order to determine the usability of the current version of the isiZulu TTS system. In particular, the LLSTI initiative is aimed at providing information to users from all walks of life, and we wanted to determine whether the TTS would be understandable to users with limited literacy and limited exposure to such technology.

The current version of the system suffers from a number of limitations: synthesis is based on a corpus of only 153 recorded sentences (the majority of which were drawn from the weather domain); tone and rhythm are handled in a precursory fashion. Loan words – which are quite common in isiZulu – are not treated distinctly from normal isiZulu words, except for the use of a distinct pronunciation dictionary for such words. We therefore also wanted to understand how much these limitations harm the naturalness of the TTS system.

2.2 Methods

Since the primary aim of LLSTI is to provide telephone-based services, one part of the pilot evaluation was designed to assess the utility of our TTS system for over-the-telephone use. Information on three subjects (a weather forecast, the disease malaria, and unemployment insurance) was scripted in English, translated into isiZulu, and synthesized with our TTS. This information was embedded in an IVR application, which provided initial instructions, and then prompted the evaluators to listen to each of the information clips. All prompts are in isiZulu; whereas the synthesized voice is male, a female voice was recorded for the rest of the IVR application. English and isiZulu versions of the IVR script are provided in Appendix A.

A questionnaire was prepared to (a) obtain biographic information of the evaluators (e.g. their age, home language, level of education); (b) test the evaluators' understanding of the matter presented in this way; and (c) query the evaluators' subjective experience of the synthesized speech. Evaluators were asked to provide their biographic information before interacting with the IVR application; understanding was tested after each information clip had been presented, and the questions on their subjective experiences were completed after termination of the IVR application. All responses were recorded on paper – where the evaluators were not sufficiently literate to perform this task, an experimenter asked them the questions verbally, and filled out the questionnaire on their behalf. Appendix B contains the questionnaire used.

Finally, a Web-based evaluation was also developed in order to gain access to a wider group of evaluators (although this evaluation would then be limited to literate, technologically sophisticated users). The same basic protocol was followed as for the telephone-based evaluation, but users now recorded their responses directly in the Web application, and their

responses were logged on the Web server. A screenshot of the Web application is presented in Appendix C.

Evaluators were canvassed in various ways, ranging from personal contacts to a company-wide e-mail solicitation. Evaluators completed the process without assistance where possible, but in cases where their level of literacy did not allow them to read the instructions and complete the questionnaire, an experimenter assisted them by reading from the questionnaire and entering their answers on the paper form. Results from the telephone-based and Web-based trials were entered into a spreadsheet, and processed to assess the performance of the TTS system.

2.3 Results

The telephony system was evaluated by a test user group consisting of twenty-three people. 40% of the test group (E1: evaluators E101 – E109) had a low literacy level. These evaluators either indicated that they could not read or write, or had significant difficulty reading the evaluation form. For this user group, typically occupations included domestic worker, gardener and laundry-worker. The remaining 60% of the test group (E0: evaluators E001 to E014) were literate, and in certain cases well educated. Occupations for this second group ranged from domestic worker and security guard to software developer and human resources practitioner. Sixteen of the telephony evaluators indicated that they spoke isiZulu as a “home language” – the remainder of the test group consisted of fluent isiZulu speakers with home languages that included Setswana, isiNdebele, SiSwati, Xitsonga and isiXhosa. A third test group (E2) evaluated the system via the Internet application.

2.3.1 E0 Evaluation

The following was noted during the E0 evaluation (literate evaluators):

- 57% of evaluators indicated that they found the synthesised voice “fairly easy” to “very easy” to understand. While the remainder of the test group found it “fairly hard” to “very hard”, no one indicated that the system was “impossible” to understand.
- Surprisingly, 50% of the evaluators found the synthesized voice “as natural as” or only “slightly less” natural than the recorded prompts used to provide the instructions. The remainder felt that the voice was either “a fair amount” or “a lot” more unnatural with only one evaluator indicating that the synthesised voice does not “sound human at all”.
- A strong correlation was observed between people’s ability to answer the test questions correctly, and the “intelligibility score” they gave the system, as shown in Table 1.

Table 1: Perception of intelligibility according to number of questions answered correctly for E0 evaluators

Number of correct answers (of 3)	Own intelligibility score (of 5)
3	4.4
2	3.3
1	2.7

- Five of the 13 evaluators (36%) answered all five questions correctly. The average number of correct answers over all evaluators was 2.1 (of 3).

- Home language isiZulu-speakers tended to answer the evaluation questions more accurately than other isiZulu-speakers (with an average of 2.3 correct answers compared to 1.8)

2.3.2 E1 Evaluation

The E1 evaluation (illiterate evaluators) produced results that differed from the above in interesting ways. During observation and through after-evaluation discussions, it became clear that the test procedure was not sufficient in two ways:

- Evaluators found it difficult to differentiate between information obtained from the system, and information they knew beforehand. For example, the weather question provided a fictional weather report and on account of the information provided, asked what the weather is “expected to be today”. A number of the evaluators provided the actual information, not the fictional information¹.
- Evaluators found it difficult to choose a negative answer – even where they knew the negative answer to be correct. For example, evaluator E106 answered the malaria question incorrectly, as he understood contact with human blood to be the way in which the mosquito contracted malaria – and as that is “bad”, chose the other answer. In the same vein, the employment related question asked whether “only an employer” should contribute to the Unemployment Insurance Fund. Even evaluators, who in subsequent questioning indicated that they understood that both the employer and employee should pay, still chose “yes” because the employer should contribute, rather than “no” (the correct answer) since the double negative required to answer the question correctly was unnatural to them.

This demonstrates the complexity involved when using a comprehension test with low literacy evaluators: an area in which we realise significant further research is required.

The subjective opinions of E1 evaluators were similar to those of E0 evaluators with regard to intelligibility (3.8 average compared to 3.6) and slightly lower with regard to naturalness (2.8 average compared to 3.2). While the objective evaluation (which does not compensate for the issues mentioned above) indicates a low level of intelligibility (1.4 average), the results of the subjective evaluation are very encouraging.

2.3.3 E2 Evaluation

The Internet evaluators (E2) produced similar trends as observed during the E0 evaluation. The Internet evaluation process has not drawn many responses, and the evaluation is being continued.

Detailed results of all three evaluations are included in Appendix D.

¹ The fact that 8 of the 9 E1 evaluators chose “sunny” rather than “cloudy” (the correct answer) testifies of South Africa’s prevalent weather conditions.

3 Conclusion

Initial experiments with the LLSTI isiZulu test system have produced encouraging results. The voice was found to be quite understandable by the majority of evaluators – both those with high levels of literacy and those with limited or no literacy. Since much of the information was outside the weather domain, which was the focus of this development, it is safe to consider this as a domain-independent result. (In fact, comprehension was best in the non-weather domains, for reasons described above.)

Clear room for improvement remains, both in terms of the characteristics of the TTS system, and in our evaluation thereof. The design of comprehension tests for users of limited literacy will require particular attention. Interestingly, the lack of attention to prosodic information did not attract any specific comments from the evaluators, suggesting that this may not be as important for understandable isiZulu TTS as had been believed.

The lessons learnt will be applied in refining our experimental approach, and in the continued development of the isiZulu TTS system.

Appendix A: Telephony evaluation dialogue

isiZulu version of IVR dialogue

Ingxoxo	
Recorded voice	<p>Usufike ohlelweni lwe CSIR lokuhumusha okubhaliwe kuyiswe enkulumweni, Text-to-speech system.</p> <p>Ukuze uvivinye loluhlelo, sicela ukuthi ulalele izindaba ezikulezihloko ezilandelayo;</p> <p>Isimo sezulu, ezempilo kanye noMshwalense wabalahlekelwe imisebenzi.</p> <p>Sicela ulandele izinkomba uma ukhetha isihloko.</p> <p>Emva kokulalela isihloko ngasinye, uzocelwa ukuthi uphendule umbuzo – sicela ugcalise impendulo yakho ephepheni olinikeziwe.</p> <p>Emva kokulalela zontathu izihloko, ungabeka phansi ucingo, bese ugcalisa ipheshana lokuvivinya loluhlelo.</p> <p>Uhlelo luyalandela.</p> <p>Isimo sezulu:</p>
Synthesized voice	<p>EGauteng izulu lizoba lihle, kodwa kwezinye izindawo kuzobanda kulale iqhwa ngasekuqaleni kosuku.</p> <p>Ngokuhamba kosuku, kulindeleke ukuthi lithi gqwa-gqwa ngamafu ezindaweni ezingase Mpumalanga.</p> <p>Ntambama, kulindeleke imikhemezelo eGoli namaphethelo.</p>
Recorded voice	<p>Manje sesicela uphendule umbuzo omayelana nesimo sezulu ephepheni lakho, bese ucinezela noma iyiphi inkinobho ocingweni lwakho uma usulungele ukuqhubekela phambili.</p> <p><Wait for key press></p> <p>Ezempilo:</p>
Synthesized voice	<p>IMalaria idalwa amagciwane ahlasele igazi lomuntu. UMyane uthola leligciwane leMalaria ngokudla igazi lomuntu onalegciwane. UMyane uyaye udlulisele leligciwane leMalaria kumuntu ngamathe awo, ngesikhathi umtinyela uzama ukudla igazi lakhe.</p>

Recorded voice	<p>Manje sicela uphendule umbuzo ngezeMpilo ephepheni lakho, bese ucidezela noma ngabe iyiphi inkinobho ocingweni lwakho uma usulungele ukuqhubekela phambili.</p> <p><Wait for key press></p> <p>UMshwalense wabalahlekelwe umsebenzi:</p>
Synthesized voice	<p>Abantu abaqashe izisebenzi zasekhaya baphoqelekile ukuba bakhokhe imali yoMshwalense wokulahlekelwa umsebenzi. Umqashi akaphoqelekile ukukhokha loMshwalense uma umsebenzi esebenza ngaphansi kwamahora angamashumi amabili nane ngenyanga. Bobabili umsebenzi kanye nomqashi kufanele bakhokhe okukodwa ekhulwini lomholo wenyanga, njalo ngenyanga.</p>
Recorded voice	<p>Manje sicela uphendule umbuzo ngezoMshwalense wabalahlekelwe umsebenzi ephepheni lakho, kanye nombuzo wokugcina.</p> <p>Siyabonga ngosizo lwakho. Usale kahle.</p>

English translation of IVR dialogue

Dialogue	
Recorded voice	<p>You have reached the CSIR Text-to-Speech system.</p> <p>In order to evaluate this system, we would like you to listen to information on three topics; Weather, Health and Unemployment insurance.</p> <p>Please follow the instructions to select each of the different topics.</p> <p>After listening to each topic, you will be asked to answer a question – please fill in the answer on the form that was handed to you.</p> <p>After listening to all three topics, you can hang up and complete the evaluation form.</p> <p>Weather information:</p>
Synthesized voice	<p>Gauteng is fine and cold with frost in places at first. Later on in the day, it is expected to become partly cloudy in the east. Light rain is expected over the Johannesburg area in the late afternoon.</p>
Recorded voice	<p>Now please answer the Weather question on your evaluation form, and press any key on your telephone when you are ready to continue</p> <p><Wait for key press></p> <p>Health information:</p>
Synthesized voice	<p>Malaria is caused by parasites that infect human red blood cells. The mosquito picks up the malaria parasites from the blood of an infected human when it feeds. The mosquito then passes the malaria parasites to another human through its salivary glands.</p>

Recorded voice	<p>Now please answer the Health question on your evaluation form, and press any key on your telephone when you are ready to continue</p> <p><Wait for key press></p> <p>Unemployment information:</p>
Synthesized voice	<p>Employers of Domestic Servants are liable to pay UIF contributions. An employer does not need to pay UIF if the employee is employed for less than 24 hours per month. Both the employer and the employee must pay one percent of the total monthly remuneration to the UIF.</p>
Recorded voice	<p>Now please answer the unemployment question on your evaluation form, and also the final questions.</p> <p>Thank you for helping us! Goodbye.</p>

Appendix B: Telephony evaluation questionnaire

isiZulu version of evaluation questionnaire

IPHEPHA LESIVIVINYO

Uyacelwa ukuthi unikezele ngeminigwane yakho. Sicela ufunde lembuzo elandelayo emva kwalokho ucindezele inkinhobo ocingweni lwakho ubese ulalela izihloko ezifundwa lapho. Uma usucedile uphendule imibuzo ngesinye nesinye isihloko.

Imininingwane ngomuntu

1. Igama nesibongo

2. Ubudala

3. Ulimi olukhulunywa ekhaya

4. Ubulili

5. Ubuhlala kuphi eminyakeni emihlanu yokuqala empilweni yakho?

6. Isikhathi sakho esiningi kuleminyaka eyishumi edlule usichithe uhlala kuphi?

Umbuzo ngesimo sezulu (yenza isiyingi empendulweni oyikhethayo)

Isimo sezulu sase Gauteng sinjani namhlanje ntambama.

(a)Libalele

(b)Liguqubele

Umbuzo ngezeMpilo (yenza isiyingi empendulweni oyikhethayo)

UMyane uthola igciwane leMalaria ngokuthinta

(a) Igazi lomuntu

(b) Amathe omuntu

Umbuzo ngezoMsebenzi (yenza isiyingi empendulweni oyikhethayo)

Umqashi uyena kuphela yini okufanele akhokhele uMshwalense wabalahlekelwe umsebenzi?

(a) Yebo

(b) Cha

Umbuzo wokugcina (yenza isiyingi empendulweni esondelene kakhulu nombono wakho)
Uma uqhathanisa izwi lomuntu wesifazane elisetshenziswe esigabeni sokuqala saloluhlelo kanye nezwi lomuntu wesilisa elisetshenziswe esigabeni sesibili saloluhlelo, ngabe izwi lomuntu wesilisa lizwakala ngendlela engajwayelekile?

- (a) Cha
- (b) Kancane
- (c) Kakhudlwana
- (d) Kakhulu impela
- (e) Akulona izwi lomuntu nhlobo.

Kube nzima kangakanani ukuzwa ukuthi izwi lomuntu wesilisa belithini?

- (a) Beyingekho nhlobo indlela yokuzwa
- (b) Kunzima kakhulu
- (c) Kunzima kancane
- (d) Kulula
- (e) Akukho nkinga

English translation of evaluation questionnaire

Zulu Text-to-speech: Evaluation form

Please fill in your personal details.

Please read the questions on each section before you dial in and listen to the various topics. Answer the questions of every section directly after listening to each section, before continuing to the next section.

Personal information:

1. Name and Surname

2. Age

3. Home Language

4. Gender

5. Where did you spend the first five years of your life?

6. Where have you spent the majority of you time over the last ten years?

Weather question: (circle the correct answer)

The weather in Gauteng later today is expected to be:

(a) Sunny (b) Cloudy

Health question: (circle the correct answer)

A mosquito becomes a Malaria-carrying mosquito from contact with

(a) Human blood (b) Human salivary glands

Employment question: (circle the correct answer)

Must only an employer pay UIF contributions?

(a) Yes (b) No

Final questions: (circle the answer that most closely matches your opinion)

Does the second part, where the male-voice is used, sound more unnatural than the female voice in the beginning?

(a) No (b) Only a little (c) A fair amount (d) A lot (e) Not human at all.

How hard was it to understand what the male voice was saying?

(a) Impossible (b) Very hard (c) Fairly hard (d) Fairly easy (e) No problem

Appendix C: Web-based evaluation

Isivivinyo sohlelo lokuhumusha okubhaliwe kuyiswe enkulumweni

Text-to-Speech System

Usufike ohlelweni lwe **CSIR lokuhumusha okubhaliwe kuyiswe enkulumweni**.

Ukuze uvivinye loluhlelo, sicela ukuthi ulalele izindaba ezikulezihloko ezilandelayo;

Isimo sezulu, ezempilo kanye noMshwalense wabalahlekelwe imisebenzi.

Sicela ulandele izinkomba uma ukhetha isihloko.

Emva kokulalela isihloko ngasinye, uzocelwa ukuthi uphendule umbuzo - sicela ugcalise impendulo yakho.

Ingxoxo nge**Simo sezulu**

Umbuzo ngesimo sezulu: Khetha impendulo yakho!
Isimo sezulu sase Gauteng sinjani namhlanje ntambama:

Ingxoxo nge**Zempilo**

Umbuzo ngezeMpilo: Khetha impendulo yakho!
UMyane uthola igciwane leMalaria ngokuthinta

Ingxoxo ngezo**Mshwalense wabalahlekelwe umsebenzi**

Umbuzo ngezoMsebenzi: Khetha impendulo yakho!

Umbuzo ngezoMsebenzi: Khetha impendulo yakho!
Umqashi uyena kuphela yini okufanele akhokhele uMshwalense walahlekelwe umsebenzi?

Umbuzo wokugcina: Khetha impendulo esondelene kakhulu nombono wakho
Uma uqhathanisa izwi lomuntu wesifazane elisetshenziswe esigabeni sokuqala saloluhlelo kanye nezwi lomuntu wesilisa elisetshenziswe esigabeni sesibili saloluhlelo, ngabe izwi lomuntu wesilisa lizwakala ngendlela engajwayelekile?

Kube nzima kangakanani ukuzwa ukuthi izwi lomuntu wesilisa belithini?

Imininingwane ngomuntu

Igama nesibongo

Ubudala

Ulimi olukhulunywa ekhaya

Ubulili

Ubuhlala kuphi eminyakeni emihlanu yokuqala empilweni yakho?

Isikhathi sakho esiningi kuleminyaka eyishumi edlule usichithe uhlala kuphi?

Internet

Appendix D: Detailed results

Key

Table 2: Literacy Level

Description	Score
Cannot read or write	0
Literate, but significant difficulty reading	1
Literate	2
Further qualification	3

Table 3: "How hard to understand?"

Option selected	Description	Score
a	Impossible	1
b	Very hard	2
c	Fairly hard	3
d	Fairly easy	4
e	No problem	5

Table 4: "Does the voice sound more unnatural?"

Option selected	Description	Score
a	No	5
b	Only a little	4
c	A fair amount	3
d	A lot	2
e	Not human at all	1

E0 Evaluation

Evaluator	Age	Home language	Gender		Place of origin	Past five years	Occupation	Literacy level	Q1	Q2	Q3	Q4	Q5
E001	31	isiZulu	owesilisa	Male	Maphumalo (kzn NorthCoast)	Durban Secunda (3); Pretoria (5)	Software developer	3	b	a	b	d	d
E002	24	isiZulu	owesifazane	Female	eThekwini	eSikoleni		2+	b	a	b	a	e
E003	39	isiZulu	sifazane	Female	Kwa-Thema	Kwa-Thema		2+	b	a	b	c	d
E004	27	isiZulu	Male	Male	ePitoli	ePitoli	Software tester	2	b	a	b	d	e
E005	29	isiZulu	Male	Male	Gembokspruit, Empumalanga	Witbank	Security guard	2	b	a	b	b	d
E006	23	isiZulu	Male	Male	eQuthu	eQuthu	Security guard	2	b	a	a	b	c
E007	28	Tswana	Male	Male	Meadowlands	Diepkloof	Security guard	2	a	a	b	b	d
E008	27	isiZulu	Silisa	Male	eDennilton eMpumalanga	lapha eGauteng ngicale eGoli ngeza efitoli	Software developer	3	a	a	b	c	c
E009	29	isiZulu	Male	Male	Mpumalanga	Johannesburg	HR Practitioner	3	b	b	b	d	b
E010	58	Tsonga	Male	Male	Lady Selborne E Pitori	Atteridgeville E Pitori		2+	a	a	b	d	d
E011	25	isNdebele	Female	Female	ePretoria	eSoshanguve		2+	a	a	b	b	d
E012	37	isiZulu	isilisa	Male	eMzumbe, South Coast, KZN	eGoli		2+	a	a	a	b	c
E013	24	isiZulu	ngowesifazana	Female	Mamelodi, Pitoli	Mamelodi, Pitoli		2+	a	a	a	a	b
E014	41	Ndebele	Umama	Female	Belfast	Pretoria	Domestic worker	2	b	b	a	e	c

E1 Evaluation

Evaluator	Home		Gender	Place of origin	Past five years	Occupation	Literacy level	Q1	Q2	Q3	Q4	Q5
	Age	language										
E101	31	isiZulu	owesifazane	Female	e Ruigtesloot	eMabopane	0	a	a	b	b	d
E102	33	isiZulu	Besifazane	Female	eMamelodi	eMamelodi	0	a	a	b	b	c
E103	27	Tsonga	owesilisa	Male	kuma-Klerksdorp	Mamelodi	1	a	a	b	a	e
E104	40	isiZulu	Female	Female	Pretoria	Pretoria	1	a	a	b	d	d
E105	42	Xhosa	Female	Female	Mount Free	Stinkwater	1	a	a	a	b	e
E106	49	isiZulu	Male	Male	Pretoria	Pretoria	1	b	b	a	d	d
E107	44	isiZulu	Umama	Female	Middelburg	Pretoria	1	a	a	a	e	c
E108	26	isiZulu	Male	Male	Natal	Pretoria	1	a	a	a	d	c
E109	51	Swazi	Umama	Female	Bushback Ridge	Pretoria	0	a	a	a	e	c

E2 Evaluation

Evaluator	Home		Gender	Place of origin	Past five years	Literacy level	Q1	Q2	Q3	Q4	Q5	
	Age	language										
E207	20+	isiZulu	Owesifazane	Female	eGoli (Johannesburg)	eKapa (Cape Town)	2+	b	a	b	d	c
E208	29	Zulu	Owesifazane	Female	Durban (eThekwini)	ePitoli	2+	b	a	b	b	e
E203	23	zulu	Owesifazane	Female	monsterlus	ethekwini nasepitoli	2+	b	a	b	b	c
E201	31	Zulu	Owesilisa	Male	Natal. Laydsmitlh	Natal Ladysmith, Empangeni, Vryheid nase Durban	2+	b	b	b	a	a
E206	26	Zulu	Owesifazane	Female	Daveyton	Durban	2+	a	a	b	a	c
E204	24	isiZulu	Owesifazane	Female	Mamelodi, Pretoria	Mamelodi, Pretoria	2+	b	a	a	d	c
E202	58	Tsonga	Owesilisa	Male	Lady Selborne e pitori	Atteridgeville e Pitori	2+	a	a	a	a	a
E205	35	isiZulu	Owesilisa	Male	EGoli	EPitoli	2+	a	b	b	a	a