

# Introduction

Languages is an emblem of identity. Preserving languages is vital in saving cultural wealth and important ancestral knowledge embedded in these languages.

Many African languages are heading for extinction and some are already extinct (like |Xam, previously spoken in the Western Cape).

Records of the language currently existing in a digitized dictionary. (Figure 1)

in a second and the second	in the cave . ! have ha I nein .
for it and the set of a contain	I heard, in Il huan this ; ,
I marked the thing that and	that, the thing to e, tos'a he
Martans. And ha Wheney	heathing die not the youtri . htwin
and search this to the later tail to	10 a man(a); 0 thui; no The har-
a year marker ghand in the the	that, I would trick and the
many and and the	feel about, She winter than,
As the second 16's When and	lay down my things . I ha tehueir .
Part & Come permit & That	20 2 jelt about, Si The han,
Constant and the second	while o

Figure 1

#### Background

The corpus of |Xam text document is complex and can't be easily input on a keyboard or small screen.

- Complex with diacritics above, below, above and/or below character(s). (Figure 2)
- No Language Model (Figure 3)

'hu i yahen ! heirya. 'gane i # huan kanh." !ho'an !huonni 'ihe'.	a	- <del>L</del>	j a	ŝ	۶۰, ۱
Element O	Elguro 2				
Figure 2	rigule s				

#### Objective

To compare the accuracy and speed of entry of QWERTY, T9, Pinyin Script and Hierarchical input methods using Xam text on mobile devices. (Figure 4-7)

qwertyuiop								
а	s	d	f	g	h	j	k	1
ୢୡ	z	x	с	v	b	n	m	$\bowtie$
?123	ļ	þ		_			•	Ψ
Figu	Figure 4. QWERTY							



×۱	N	e r		t	у	ui			
a	s	d	g	h	nk lo		о		
Ŷ	р	с	b	n	n	n	$\langle \times \rangle$		
<b>\$</b>	?	123	_	_	• •		Ļ		
Figure 5. T9									
	_	_	_	_	_				



Figure 7. HIERARCHICAL

# Xamobile: Comparing mobile text input methods for historical African languages Sunkanmi Olaleye **Department of Computer Science, University of Cape Town.**

### Hypotheses

Text of the language should be preserved and mobile text entry could help. Figure 8 shows images of Xam text with diacritics from the notebooks and their transcriptions. Xam consists of about 35000 unique characters. These

characters consist of single/double characters with diacritics below or above or above and below with the inclusion of special clicks and symbols.

Image	Transcription		
Thu i yaken thei-ya.	!kǘi yaken̆ !kuı̈́-yä́.		
'gave à # heran kank."	!gau≀ë #kua'n kaŭkí		
Ko'an ikuonni like.	$ ! k \circ \ddot{\ddot{a}} h k \breve{u} onn \breve{i} ! \breve{u} h \breve{e} . $		

Figure 8

# **Research Questions**

- 1. How do the XWERTY, T9, Pinyin Script and Hierarchical input methods compare in terms of accuracy for |Xam text?
- 2. How do the XWERTY, T9, Pinyin Script and Hierarchical input methods compare in terms of Speed of entry for |Xam text?

# Methodology

Baseline Study. Figure 9 shows the designed |Xam font required for entry of single/double characters with diacritics.

- User Centered Design
- Prototyping. (Figure 10)
- Development (Figure 11)

# Testing and Evaluation (Figure 12)

	10500	1050.5	10500	10505		1054.0	10544	1054.0	10540	1054.4
uni250A	uni250B	uni250C	uni250D	uni250E	uni250F	uni2510	uni2511	uni2512	uni2513	uni2514
		<u>ل</u>	<u> </u>	~	<u> </u>	1	r		~	$\mathbf{\hat{\mathbf{x}}}$
nn		an	a			k	n	a	a	Α
	υu		u	u	u	ų.		u	u	1.
10510	10514	10510	1051.0		10545	10515	10500	105.01		10500
uni2519	uni251A	uni2518	uni251C	uni251D	uni251E	uni251F	uni2520	uni2521	uni2522	uni2523
<u>ک</u>	Ň	ટ	X	ິ	X	U U		1	É.	7
a	Α	a l	Α	a	A	a		Α		Α
u	1	u	1	ч	1.	ų	<u> </u>			
	105.00	10504	10500		10500	in lanar	in lanar	10520	10524	
uni2528	uni2529	uni252A	UNI252B	uni252C	uni252D	uni252E	uni252F	uni2530	uni2531	uni2532
Ê	~	Ê	~	Ê	~	Ê	^	f	<u>~</u>	т
	ρ		ρ		ρ		ρ			
						• •	÷	-	•	÷
uni2537	uni2538	uni2530	uni253A	uni2538	uni253C	uni253D	upi253E	upi253E	upi2540	uni25/11
um2557	um2556	um2555	UTI255A	unizooo		um2550	~	umzoon	01112.540	0112341
Ľ.	Â	~	Ô	<u>``</u>	Ô	<u>~</u>	Ô	ã	Ô	<u>^</u>
0		0		0		0		0		0
X				0					Ý	Ŷ
Elaur	0 0									
iyur										









# Development

After several design iteration cycle with our focus group, we feel confident that it fits requirement and everything that has been learned through the iterated steps of prototyping and evaluation are integrated to produce the final prototype to be used for

#### Apparatus

Xam Line Text (Source: gold standard data used in AHR and TBL)

- Android Touchscreen mobile phones
- 4 Prototype Input Methods

AHR : Automatic Handwriting Recognition [Williams and Suleman 2011] TBL : Transcribe Bleek & Lloyd [Ngoni and Suleman 2013]

text

Center of Excellence in Broadband networks and applications

VC74D Department of Computer Science, University of Cape Town, Rondebosch 7700

Figure 12



**Results and Discussions** 

$$WPM = \frac{|T| - 1}{S} * \frac{60}{5}$$

$$KSPC = \frac{|InputScream|}{|TranscribedText|}$$

$$Total \ Error \ Rate = \frac{INF + IF}{C + INF + IF} * 100\%$$

$$ASD \ Error \ Rate = \frac{MSD(A, B)}{\overline{S_o}} * 100\%$$

•WPM is the Word Per Minute metric for measuring speed of text

- entry. |T| is the length of transcribed |Xam text
- •S is the duration in seconds
- •Note that the constants 60 and 5 represent conversion metric for *minute and average word length of Xam text.*
- •KSPC is the keystrokes Per Character
- InputScream is the total number key-press for the presented Xam text
- •|TranscribeText| is the total number key-press required using a
- particular input method for the presented Xam text
- INF represents the total number of errors made but not corrected in the Xam text
- IF represents the total number of errors made and corrected in the Xam text
- C represents the total number of correct keystrokes in the Xam

# Conclusions

Quantitative data would be collected to calculate WPM, Total Error Rate and MSD metrics for the purpose of comparison

# Acknowledgement

# Contact information

Supervised by: A/Prof. Hussein Suleman