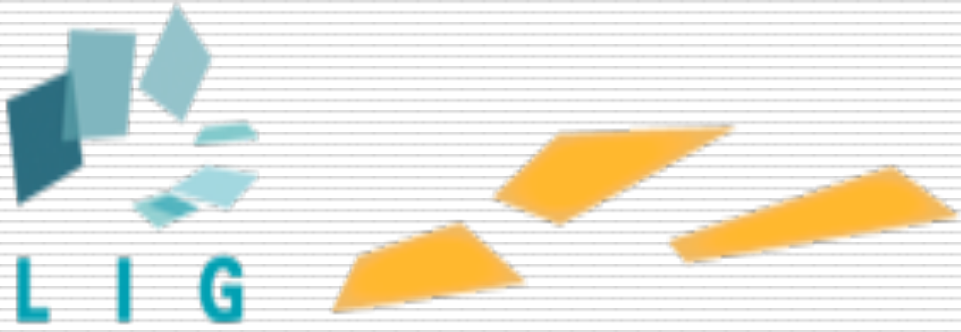
Amharic Speech Recognition for Speech Translation

Michael Melese Woldeyohannis, AAU, Addis Ababa, Ethiopia

Laurent Besacier, LIG Laboratory, Univ. Grenoble Alpes, France

Million Meshesha, AAU, Addis Ababa, Ethiopia



Context : ALFFA project (1/2)

- African Languages and Information Technologies
- Address under-resourced languages from Africa
- Focus on West Africa
  - Hausa, Wolof, Fulfulde, Zarma, Bambara
- 2 East African languages
  - **Amharic**, Swahili
- Data collection methodology
- ASR (speech-to-text) and TTS (text-to-speech)
- French Partners : LIG (Grenoble), LIA (Avignon), DDL (Lyon), Voxygen (Lannion), http://alffa.imag.fr

Context : ALFFA project (2/2)

- This paper focuses on Amharic ASR for a specific domain
- Four languages covered so far in ASR
  - Hausa, Wolof, Amharic, Swahili

Task	WER %
Swahili Boadcast News	20.7
Hausa Read Speech	10.0
Amharic Read Speech	8.7
Wolof Read Speech (under dev.)	27.2

- Data and Kaldi [1] scripts released on *Github*
- https://github.com/besacier/ALFFA_PUBLIC

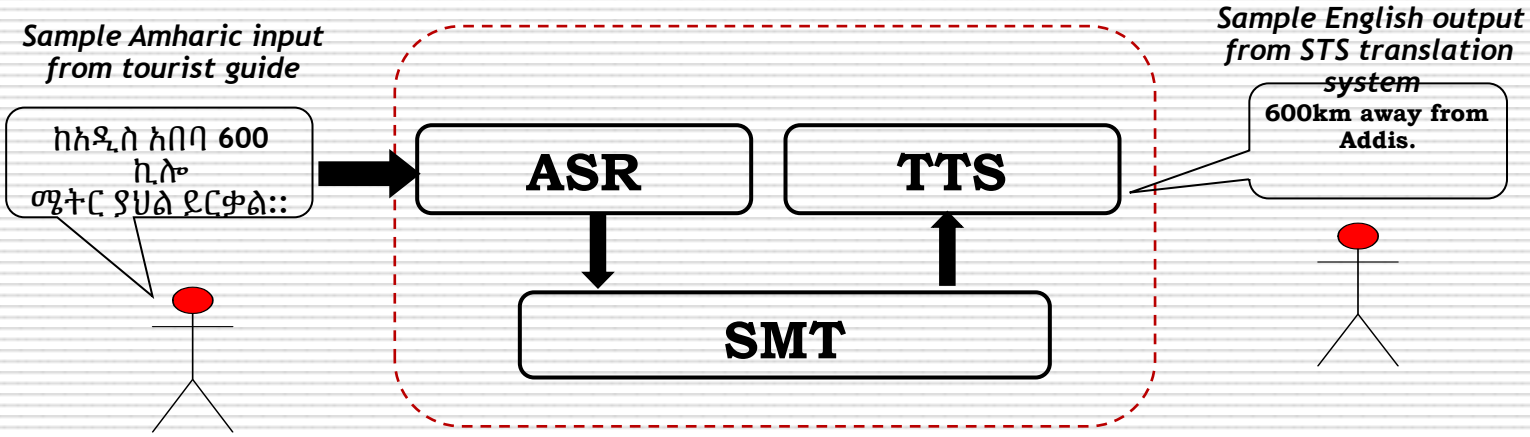
[1] D. Povey & al., "The kaldi speech recognition toolkit," IEEE ASRU 2011.

Amharic ASR and SLT for tourism (1/2)

- Ethiopia has much to offer for international tourist
  - It is a land of natural contrasts, ranging from the peaks of the rugged Semien mountains to the depths of the Danakil depression, which is one of the lowest points on earth more than 400 feet below sea level.
- According to UNWTO and World Bank report, the number of tourist grows 14% every 5 year to visit different locations in Ethiopia
  - Tourist attraction including world heritages, which are registered as Ethiopian tourist attractions by UNESCO.

Amharic ASR and SLT for tourism (2/2)

- This paper focuses on ASR
- Start from the BTEC (*Basic Travel Expression Corpus*) initially available in English
- BTEC translated into Amharic by AAU



a need to develop a speech translation system so that tourists can effectively communicate with the tourist guide regardless of the language that they speak.

Amharic language (1/2)

- Amharic is the 2nd largest spoken Semitic languages in the world after Arabic among 89 registered languages.
- Unlike other Semitic languages, such as Arabic and Hebrew, Amharic */ʾämarəñña/* script uses a writing system called *fidel /fidälə/*
- Syllable structure
  - Questions the choice of the best unit for acoustic modelling in ASR
- Rich morphology
  - Questions the choice of the best unit for language modelling in ASR

Amharic language (2/2)

Category	Character set	Order	Total
Core characters	33	7	231
labiovelar	4	5	20
labialized	18	1	18
labiodental	1	7	7
		Total	276

Distribution of Amharic character set

	Order						
	1st	2nd	3rd	4th	5th	6th	7th
	ə	u	i	a	e	ɨ	o
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
h	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
m	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
s	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ
r	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
s	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ

Unnormalized sample Amharic core character

Normalizing to distinct sound rather than with orthographic representation

	Order						
	1st	2nd	3rd	4th	5th	6th	7th
	ə	u	i	a	e	ɨ	o
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
m	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
r	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
s	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ

Normalized sample Amharic core character

Related Works (Amharic Speech Translation)

	Author	Problem Solved	Performance	Research Direction
ASR	**Solomon Birhanu (2001)**	Investigate the Consonant-Vowel syllable recognition for the Amharic language	Recognition accuracy of 87.68 for Speaker Dependent and 72.75 Speaker independent	towards speaker independent recognition of speech and tuning the model to diverse environment including.
	Solomon Teferra (2005)	Develop a large vocabulary, speaker independent continuous Amharic speech recognition using syllable and triphone.	Recognition accuracy of 90.43 % for Syllable based and 91.31% for Tri-phone.	Improving the performance of syllable and triphone ASR for Large Vocabulary.
	Tachbelie, et. al, (2014)	Selecting acoustic, lexical and language modeling units for Amharic ASR	3% absolute WER reduction as a result of using syllable acoustic units in morpheme-based LM.	syllable AM in morpheme-based speech recognition to be tested for other morphologically rich language
SMT	**Sisay Adugna (2009)**	English-Afaan Oromo machine translation system to assist professional translators.	BLEU Score of 17.74%	possibility of exploring for other local language to make the information available in all local language.
	Mulu Gebreegziabher, et. al, (2012)	Preliminary experiments on English-Amharic statistical machine translation	BLEU score result is 35.32	The experiment have been extended to get a better result out of translation.
	Mulu Gebreegziabher, et. al, (2015)	Phoneme-based English-Amharic SMT	BLEU score of 37.53 for the phoneme-based EASMT system	Further improvement of English-Amharic SMT though different technique
TTS	**Henock Leulseged (2003)**	Concatenative Amharic TTS synthesis for Amharic Language	88% using Diphone and 75% for syllable based recognition	Overcome the problems of germinated sounds for syllable and diphone based synthesis.
	Sebsibe et. al, (2004)	Unit Selection Voice For Amharic Using Festvox	Perceptual evaluation of the synthesizer showed that the quality of the voice is good	Improving by proper selection of unit and optimal corpus which covers all basic units and variations.

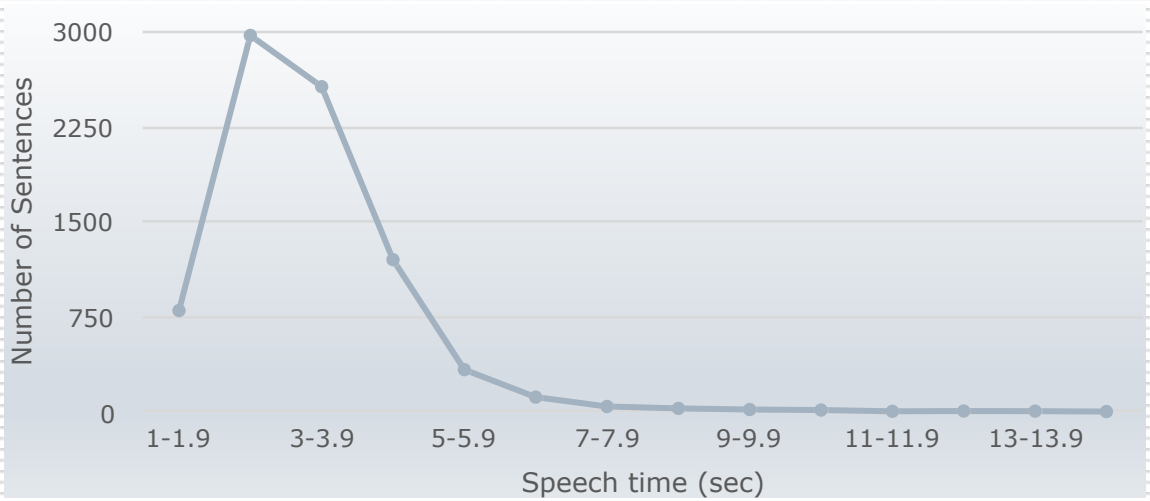
Speech corpus preparation

- A 20h Amharic read speech prepared by Tachbelie et al, (2014) is used for **training** which is available at https://github.com/besacier/ALFFA_PUBLIC/tree/master/ASR/AMHARIC
- **Testing** data from initial English BTEC (Kessler, 2010).
  - English corpus is translated to Amharic to prepare parallel Amharic-English BTEC using a bilingual speaker.
  - Amharic speech data is recorded using *Lig-Aikuma* under normal office environment by 8 native Amharic speakers (4 male and 4 female) with different age range.

Speech corpus details

	Age and Gender			
	Male		Female	
	18-30	31-50	18-30	31-50
Number of Utterances	1000	1112	1000	1000
	1000	1000	1000	1000
Total	**2000**	**2112**	**2000**	**2000**

Distribution of utterance



Speech length vs sentence distribution

A total of **8112 sentences** with a length ranging from 1 to 28 word length have been recorded.

A total of **7.43h read speech corpus** collected with an average speech time of 3.3s. Out of these utterance 98.54% of the speech data fall below 7s

Text corpus

- A text corpus collected from in-domain and out-domain data separately
  - **Out-domain** data consist of 219,631 sentences (4M tokens) of 319,858 types.
  - **In-domain** data contains 22,616 sentences (114k tokens) of 17,694 types. Subset of BTEC not used in speech recordings.
- A total of 242,247 sentences has been used to train 3-gram language model

Experiments

- In our experiments, we use Kaldi for ASR, SRILM for LM and Morfessor for unsupervised segmentation of word into sub-word unit
- **Words and Morphemes** are used as **language model** and **phone and syllable** as **acoustic** unit.

		Phoneme	Syllable
Morpheme based LM	CRA	**89.1**	85.5
	MRA	**80.9**	75.8
	WRA	**80.6**	75.8
	SRA	**49.3**	43.4
Word based LM	CRA	70.1	69.7
	MRA	52.3	50.9
	WRA	56.0	54.7
	SRA	13.2	13.2

ASR performance (character, morph, word, sentence accuracies)

Discussion

- Surprisingly, phoneme units looks better than syllable units for acoustic modeling
  - The trend was different on a large vocabulary speech recognition task (Tachbelie & al. 2014)
  - Needs more investigation

- The morpheme is a better unit than word for language modeling
  - Confirms previous findings (less OOVs)
  - Word reconstruction from morphemes not trivial

		Phoneme	Syllable
Morpheme based LM	CRA	**89.1**	85.5
	MRA	**80.9**	75.8
	WRA	**80.6**	75.8
	SRA	**49.3**	43.4
Word based LM	CRA	70.1	69.7
	MRA	52.3	50.9
	WRA	56.0	54.7
	SRA	13.2	13.2

ASR performance (character, morph, word, sentence accuracies)

Conclusion and future work

- Our experiments show that the best recognition results achieved at morpheme-based LM with phoneme-based AM. Correspondingly vocabulary gain and semi-supervised segmentation need further investigation to improve Amharic ASR.
- BTEC (Tourism domain) corpus collected for Amharic (text and speech)
- Future design of Amharic-English and English-Amharic speech translation

አመሰግናለሁ
Thank you!