Low resource multilingual TTS in NST: Blizzard & NewsHack

**Background**

Voices: easy. Languages: hard!

Why is it hard?

Speech data:
- Recordings

Expert-compiled resources & annotated data:
- Lexicon
- Phone set
- Part-of-speech tags

TTS database → ‘found’ data

Lexicon → letter-based TTS

- Assume UIU-Hi input, and rely heavily on it for:
- Tokenisation: (p(L) | p(N) | p(M))
- ASCII letter representations:

Phone set & POS tags → vector space model

A low rank approximation of a cooccurrence matrix of letters, words, etc. can be used in place of phonetic categories, POS tags, etc.

Intelligibility results

How many times less intelligible than natural speech is synthetic speech? (2011 Blizzard Festival benchmark: 1.47)

**TTS in 6 Indian languages**

Blizzard 2013: strict grapheme-based conversion. Each unicode character was considered an individual letter, relying on quinphone context in producing correct pronunciation.

Blizzard 2014: added naive support for Indic script family, using unicode names of the characters:

Blizzard 2015: added trainable inherent vowel deletion module for use in Indo-Aryan group (Hindi, Bengali and Marathi)

Intelligibility: WER of listeners’ transcriptions of SUS:

<table>
<thead>
<tr>
<th>Language</th>
<th>Bengali</th>
<th>Hindi</th>
<th>Malayalam</th>
<th>Marathi</th>
<th>Malayalam</th>
<th>Tamil</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEGIJ</td>
<td>BDFHJ</td>
<td>BCEGIJ</td>
<td>CEJH</td>
<td>EFG</td>
<td>BCDPHIJCG</td>
<td>EFGHJ</td>
</tr>
</tbody>
</table>

Swahili TTS

The BBC are piloting a tool based on computer-assisted translation and TTS for the rapid reversioning of content in different languages.

Bottleneck: availability of technology in languages of interest, e.g. Swahili

We demonstrated our technology - machine translation from subtitles, TTS, and automated audio mixing - applied to video reversioning.

Demo Swahili voice prepared using our toolkit and web-scraped data at BBC NewsHack event.

**Swahili TTS**

The BBC are piloting a tool based on computer-assisted translation and TTS for the rapid reversioning of content in different languages.

Bottleneck: availability of technology in languages of interest, e.g. Swahili

We demonstrated our technology - machine translation from subtitles, TTS, and automated audio mixing - applied to video reversioning.

Demo Swahili voice prepared using our toolkit and web-scraped data at BBC NewsHack event.

**Intelligibility results**

How many times less intelligible than natural speech is synthetic speech? (2011 Blizzard Festival benchmark: 1.47)

**Swahili TTS**

The BBC are piloting a tool based on computer-assisted translation and TTS for the rapid reversioning of content in different languages.

Bottleneck: availability of technology in languages of interest, e.g. Swahili

We demonstrated our technology - machine translation from subtitles, TTS, and automated audio mixing - applied to video reversioning.

Demo Swahili voice prepared using our toolkit and web-scraped data at BBC NewsHack event.

**Swahili TTS**

The BBC are piloting a tool based on computer-assisted translation and TTS for the rapid reversioning of content in different languages.

Bottleneck: availability of technology in languages of interest, e.g. Swahili

We demonstrated our technology - machine translation from subtitles, TTS, and automated audio mixing - applied to video reversioning.

Demo Swahili voice prepared using our toolkit and web-scraped data at BBC NewsHack event.