

Graphemes sound like a really bad idea

- mapping from phone string to acoustics has enough problems
- mapping from letter string must be even worse
- letter-to-sound is hard (especially for English)

Graphemes: why would we want to use them?

- the ultimate engineering solution?
- no lexicon required (good news for some languages)
- let the Gaussians do all the work
- easy to deal with previously-unseen words (OOVs)
 - spoken term detection
 - tolerant to mis-spellings (in training data transcription, or in query term)

Graphemes: what makes us think they would work?

- in phone-based system, we need context-dependent models:
 - words \rightarrow phonemes \rightarrow triphones \rightarrow states
- just change some of the layers:
 - words \rightarrow letters \rightarrow tri-letters(?) \rightarrow states
- it's just a different way to arrive at a state sequence

Graphemes: existing multilingual results

- Kanthak and Ney use tied-state context-dependent grapheme models
- Dutch, French, Italian, German with average of 6 hours of speech per language
- results
 - monlingual systems: graphemes are basically as good as phonemes
 - multilingual system (1) with shared models for all shared letters (monolingual models for ß, ü, ê, ç, etc)
 - a little worse than monolingual systems
 - multilingual system (2) allows questions about language during HMM clustering
 - better than (1) and not much worse than monolingual systems (actually better for German)
 - they plan to use GlobalPhone in future...

Graphemes: refinement using multigrams

- multi-letter sub-word units mapping on to state sequences
 - words \rightarrow letter n-grams \rightarrow state m-grams \rightarrow states
 - e.g. “recognition”
 - \rightarrow “re” + “cog” + “ni” + “tion”
 - \rightarrow state sequence for “re” + state sequence for “cog” + ...

Hidden Model Sequence HMMs (HMS-HMM)

- Hain's thesis
- allows arbitrary mapping from phone sequence to HMM sequence
- mapping is jointly optimised along with HMM parameters
- one implementation of the mapping uses multigrams

HMS-HMM extended to multilingual case

- language-specific components:
 - words to phones
 - probabilistic mapping from phone sequence to model sequence
- language-independent (shared) components:
 - the set of HMMs
- implementation:
 - tie the HMM parameters across languages
 - jointly optimise all the phone-to-HMM mappings on a multilingual corpus (i.e. accumulate on each corpus, pool accumulators, joint M step to update the tied HMM parameters)
- the HMMs don't have to be of phones?
- don't have to have a phone layer at all (e.g. could use graphemes)?