How Should a Speech Recognizer Work?

Scharenborg, O. Norris, D. ten Bosch, L. McQueen, J.M.

Presented By

Neal Audenaert

Agenda

- Introduction
- Computational Analysis
- The SpeM System
- Evaluation
- Discussion

Agenda

- Introduction
- Computational Analysis
- The SpeM System
- Evaluation
- Discussion

Introduction

- Human Speech Recognition
  - Models human performance in word recognition
  - Good simulation of observed behavior
- Automated Speech Recognition
  - Engineering models for word recognition
  - “Accurate” recognition of “real” speech data

They’re a Perfect Match . . .

So why can’t they get along?

Human Speech Recognition

- Focus on specific issues
  - Acoustic variability [Elman & McClelland, 1986] [Stevens 2002]
  - Lexical segmentation [Norris, et al. 1997]
  - Temporal Constraints [Marslen-Wilson, 1987] [Marslen-Wilson, & Welsh 1978]
- Success in simulating empirical data
- Piecemeal approach
- Difficult to ascertain plausibility

Automatic Speech Recognition

- Success measured by accuracy
  - Unencumbered by psychological plausibility
- Task oriented
- Not mapped to observable human behavior
- Consequently. . .
  - “Any practical ASR system is unlikely to be a candidate for a psychological theory.”

The Diagnosis

Need to focus on shared goals

What and Why, not How

- Levels of information processing [Marr, 1982]
  - Computational
  - Algorithmic
  - Implementational

Agenda

- Introduction
- Computational Analysis
- The SpeM System
- Evaluation
- Discussion

Computational Analysis

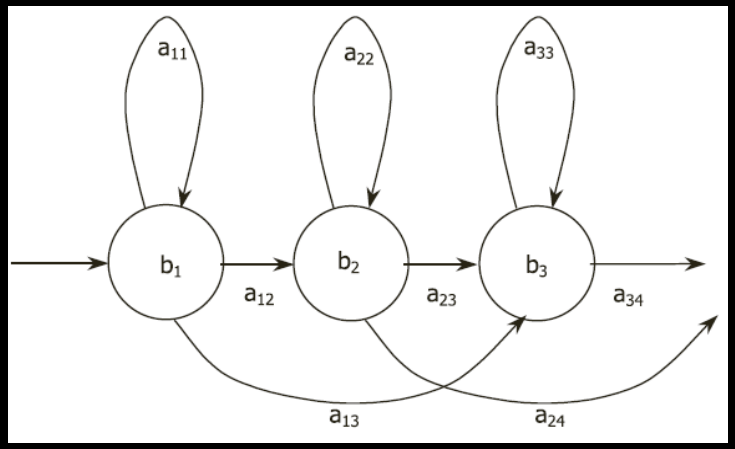
- Prelexical and lexical levels
- Cascaded prelexical level
- Multiple activation and evaluation of words
- Continuous speech recognition
- Lexical segmentation cues
- No feedback (lexical to prelexical)

Prelexical and Lexical Levels

- Invariance Problem [Perkell & Klatt, 1986]
  - Recognizers must account for variability in acoustic realization of phonemes and words
- Episodic (i.e. lexical) Models
  - Each phoneme is associated with multiple representations
  - Cannot support robust generalization to new words by humans
- HSR systems assume a prelexical component
- Major research topic in HSR [McQueen, 2005]
  - Details of the prelexical level are unknown

Prelexical Levels in ASR

- Build sub-word statistical models
- Extract features from acoustic signal
- Build and train a statistical acoustic model for each recognition unit (phoneme)
  - Hidden Markov Models
  - Artificial Neural Network



COME BACK TO ME

Cascaded Prelexical Level

- Continuous rather than discrete prelexical level
  - Lexical processing modulated by fine-grained acoustic-phonetic information [Andruski, et al. 1994] [Davis, et al., 2002] [others]
  - Lexical processing is continuous and incremental [Allopenna et al., 1998] [Zwitserlood, 1989]
- Cascaded information flow
  - Allows efficient recognition [McQueen, et al., 2003]
  - Contextual information affects lexical selection

COME BACK TO ME

Multiple Activation & Evaluation

- Psycholinguistic data
  - Multiple candidate words are activated [Allopenna, et al., 1998] [Gow & Gordon, 1995] [Tabossi, et al. 1995] [Zwitserlood, 1989]
  - Evaluation of alternatives [Cluff & Luce, 1990, McQueen et al. 1994] [Norris, et al. 1995] [others]
- Competition Mechanism
- Parallel Evaluation

COME BACK TO ME

Continuous Speech Recognition

Speechiscomparabletohandwrittentextswithoutspaces

- Parse continuous input
  - Both ASR and HSR model this as search problem
  - Extension of isolated word recognition
- HSR
  - Best parse is generated by lexical competition
- ASR
  - Score different paths through lexical lattice

Lexical Segmentation Cues

- Acoustic cues for word segments (unreliable)
  - Phonotactic constraints [McQueen, 1998]
  - Prosody [Cutler & Noris, 1998]
  - Acoustic and allophonic cues [Church, 1987]
  - Pauses [Norris, et al. 1997]
- Possible Word Constraint (PWC)
- Help with out-of-vocabulary words
- Not used in most ASR

Feedback from Lexical to Prelexical

- Subject of much debate in HSR
- Not Used in ASR
  - Unified search considers all levels simultaneously
- Appears to be no function for feedback
  - Except for “lexically-guided” learning

Computational Problems

- Invariance
- Real-time processing
- Lexical embedding
- Segmentation
- Out-of-vocabulary

Agenda

- Introduction
- Computational Analysis
- The SpeM System
- Evaluation
- Discussion

Overview & Objectives

- Concrete demonstration of parallels between ASR and HSR
- Build an HSR system using ASR components
  - Treat key components in the HSR system black boxes
  - Use ASR based implementations to approximate those components

Overview & Objectives

- Three Components
  - Automatic Phone Recognizer (APR)
  - Lexical Search Module
  - Alternative Path Evaluator

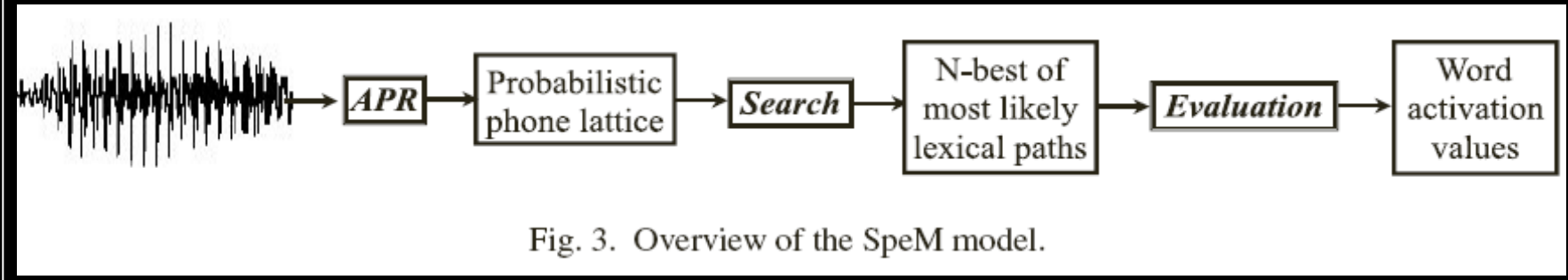
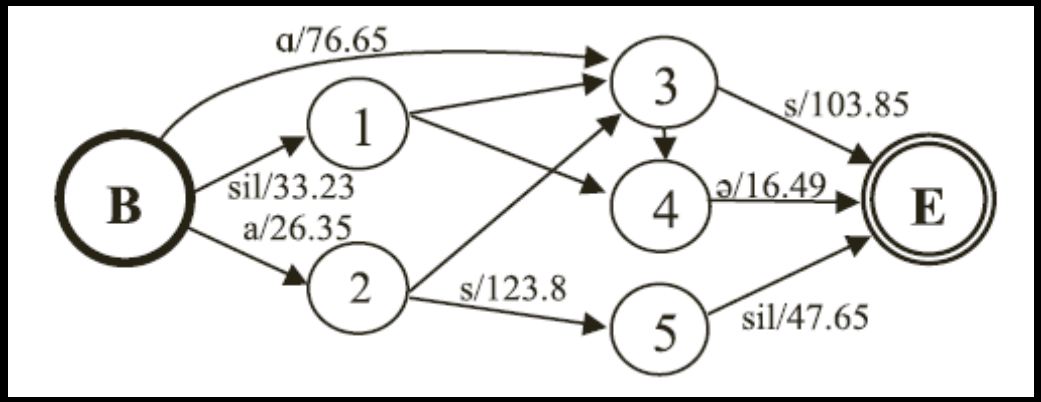


Fig. 3. Overview of the SpeM model.

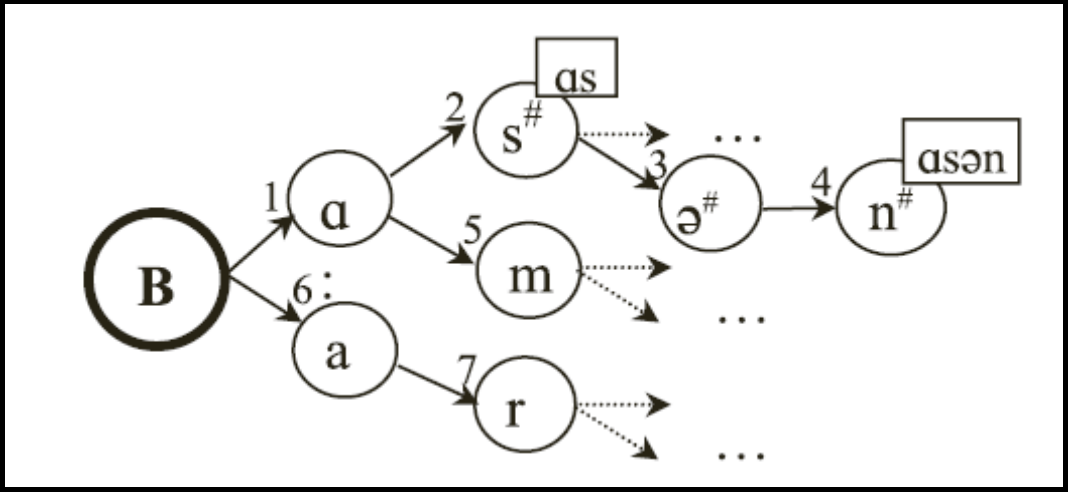
Prelexical & Lexical Levels

- Prelexical: APR generates a phone lattice
  - Weighted probabilistic lattice
  - No lexical knowledge



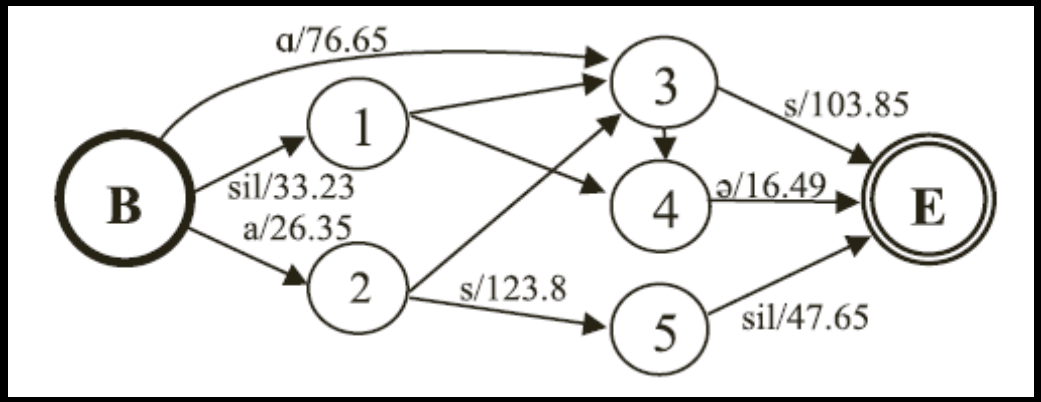
Prelexical & Lexical Levels

- Lexical Level: Search & Evaluation
  - Lexicon is represented as a tree
  - Used to find best path (Search)
  - Paths are evaluated (Evaluation)



Cascaded Prelexical Levels

- Cascaded, continuous output from ASR
- Probabilistic vs. Categorical



Multiple Activation

- Search Space
  - Product of lexical tree and probabilistic phone lattice
- Time synchronous Viterbi search
  - Hypothesized words are assigned a cost
  - Cost corresponds to the degree of match of this word to input
  - Words of dropped as cost becomes to high
- Thus: Multiple activation

Cost Functions

- Bottom-up Acoustic Cost
  - Negative log likelihood calculated by APR
- Symbolic phone matching cost
  - Accounts for substitution, insertion, and deletion
- PWC cost
- History Cost
- Word entrance penalty

Pruning

- Number of nodes
  - Maximum number of search space nodes
- Local score pruning
  - Only create a new search space if it costs less than the total cost of the best path up to this point
- No duplicate paths

Segmentation of Continuous Speech

- Implementation of possible word constraint (PWC)
  - Based on Shortlist model
- Stretch of speech edge of candidate word and the likely edge of a word boundary, that parse is penalized.

Competition & Word Activation

- Search model outputs *N* best alternative path
  - Scores are not normalized to each other
  - Computes path-based rather than word based scores
- Inadequate for HSR output
  - Goal is not the “correct” answer
  - continuous measure for ease of human recognition

Fig. 3. Overview of the SpeM model.

Competition & Word Activation

- A simplified measure for word activation
  - Bottom-up evidence
  - Probability of path the word lies on
- Activation computed using
  - Probability of the word itself
  - The best path containing the word
  - Normalized by (word + path probability) / ∏ path probablities
- Close parallels with ASR confidence measures

Competition & Word Activation

- Joint Metric of
  - Goodness of fit of a word to a stretch of input
  - Goodness of fit of the path of the word to the complete input
- Estimate changes over time as input unfolds

No Feedback

- Intermediate phonemic representation is not altered

COME BACK TO ME

Agenda

- Introduction
- Computational Analysis
- The SpeM System
- Evaluation
- Discussion

Experiment 1: Word Recognition

- Questions
  - Does it work?
  - What is the impact of probabilistic vs categorical APR
  - What is the impact of insertion, deletion, substitution
- Method:
  - Dutch Directory Assistance Corpus
  - Real speech data
  - Isolated words (Dutch city names)

Experiment 1: Results

Model	Accuracy (%)
Shortlist	32.5
SpeM-cat	36.0
SpeM + I/D/S	72.1
SpeM – I/D	70.3
SpeM – I/D/S	64.3

Experiment 1: Results

- DP search in SpeM > lexical lookup in Shortlist
- Probabilistic prelexical module >> categorical
- Insertion, deletion, substitution help
  - substitutions are most important
- Surprise, it works!
  - Not as good as state of the art ASR
  - Twice the performance of HSR

Experiment 2: Temporarily Lexically Ambiguous Input

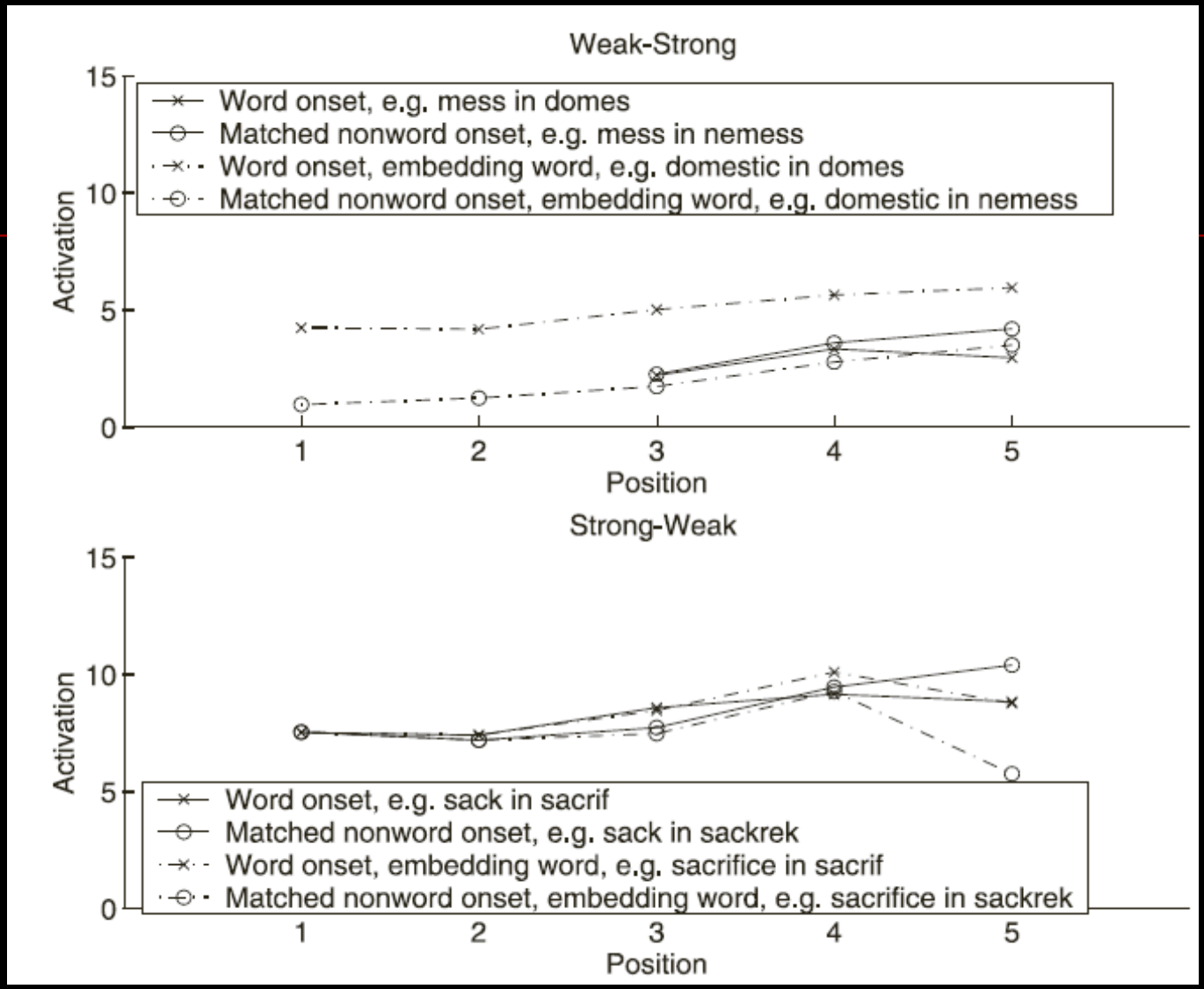
- Test ability to recognize continuous speech
  - Ship Inquiry
- Results
  - Inquiry is hardly ever substituted for another word
  - APR seems to be a limiting factor

Experiment 3: Lexical Competition

- Speech fragments with multiple candidates
- Methodology:

	Words Embedded as Second Syllable of WS Words			Words Embedded as First Syllable of SW Words		
	Stimulus	Target	Embedding Word	Stimulus	Target	Embedding Word
Word onset	do*mes*	*mess*	domestic	*sac*rif	*sack*	sacrifice
Nonword onset	ne*mess*	*mess*	—	*sack*rek	*sack*	—

Note. WS = weak–strong; SW = strong–weak.


Weak-Strong
Activation
Position
Word onset, e.g. mess in domes
Matched nonword onset, e.g. mess in nemess
Word onset, embedding word, e.g. domestic in domes
Matched nonword onset, embedding word, e.g. domestic in nemess
Strong-Weak
Activation
Position
Word onset, e.g. sack in sacrif
Matched nonword onset, e.g. sack in sackrek
Word onset, embedding word, e.g. sacrifice in sacrif
Matched nonword onset, embedding word, e.g. sacrifice in sackrek

Experiment 4: PWC & Segmentation

- More examination of the segmentation problem\
- Methodology:

	Monosyllabic Words		Bisyllabic Words	
Residue	Preceding Context	Following Context	Preceding Context	Following Context
Impossible	*f*egg	sea*sh*	*f*apple	sugar*th*
Possible	*maff*egg	sea*shub*	*vuff*apple	sugar*thim*

Experiment 4: Results

- Without PWC
  - SpeM failed to correspond to observed human recognition
- With PWC
  - SpeM corresponded to observed human recognition
- Data (??)

	Monosyllabic Words (%)						Bisyllabic Words (%)					
	Preceding Context			Following Context			Preceding Context			Following Context		
Residue	H	1B	10B	H	1B	10B	H	1B	10B	H	1B	10B
Impossible	52	79	29	39	85	6	18	98	40	38	96	17
Possible	57	77	19	28	69	10	14	90	40	17	92	4

Agenda

- Introduction
- Computational Analysis
- The SpeM System
- Evaluation
- Discussion

Discussion

- Attempted to bridge the HSR – ASR gap
- Computation analysis of word recognition process
- SpeM illustrates parallels between HSR and ASR at the computation level
- Key Feature
  - Separation of prelexical and lexical levels
  - Probabilistic (vs. categorical) prelexical level
  - DP lexical search algorith
  - Insertion deletions and substitution