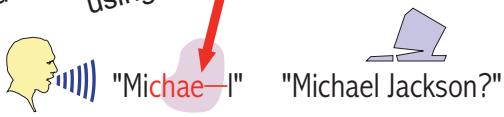
Speech Completion

Concept

You can input uncertain phrases using filled pauses!



New Direction of Speech Interface

❑ Exploit nonverbal speech information

- Current speech-input interfaces have not fully exploited the potential of speech

❑ Why human-human speech communication is comfortable?

➡ A listener sometimes helps a speaker when the speaker utters incomplete info.

- A speaker cannot remember the last part of a phrase "Michael Jackson" and hesitates

➡ A listener can help the speaker recall it

Filling in the rest of a fragment
||
Completion

Previous "Completion" Interfaces

❑ Completion in text interfaces

- Text completion has been widely used

Text editors (Emacs), UNIX shells (tcsh/bash)

➡ Provide functions completing the names of files and commands

"Completion-trigger key"  TAB key

WWW Browser

➡ Automatic completion of URLs

❑ Completion in speech-input interfaces?

- Effective functions have not been proposed

➡ There has been no way to trigger them during natural speech input

Speech Completion

❑ What is speech completion?

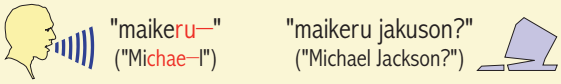
- Help a user enter an uncertain word/phrase by completing the missing part of a partially uttered fragment
- Benefits
  - A user can easily recall uncertain phrases
  - Less labor is needed to input a long phrase
  - Not forced to utter the entire content carefully, as is required by the current recognizers

❑ How to invoke the completion function?

- What is good completion-trigger key for speech?

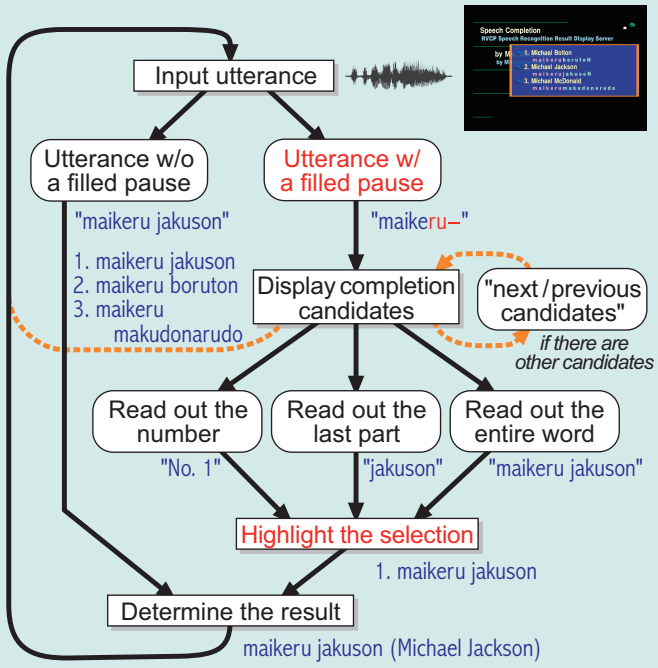
➡ Filled Pause 

- Natural hesitation that indicates a user is having trouble thinking of (recalling) a subsequent word
- Can invoke the completion function intentionally
  - Frequently used in the same way in *Japanese* conversation



Speech Interface w/ Completion

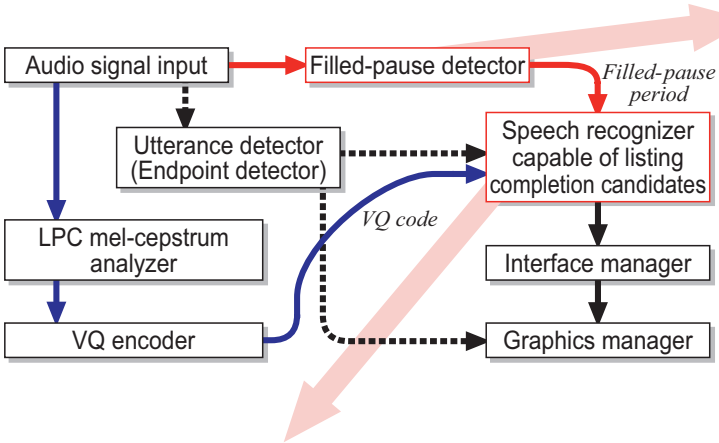
❑ Flowchart (word/phrase-level completion)



Word/phrase: Word registered in the system vocaburary
Filled pause: Lengthening of a vowel during hesitation

On-demand Completion Assistance Using Filled Pauses for Speech Input Interfaces

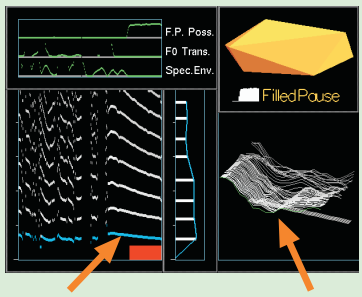
Implementation



Filled-Pause Detector

❑ Detect the beginning of each filled pause

- Real-time filled-pause (FP) detection method [Goto *et al.* 1999]
  - Independent of vocabulary and language
  - Detect a lengthened vowel in any word
- Bottom-up acoustical analysis
  - Two features of filled pause (FP)



Small pitch transition

Small spectral envelope deformation

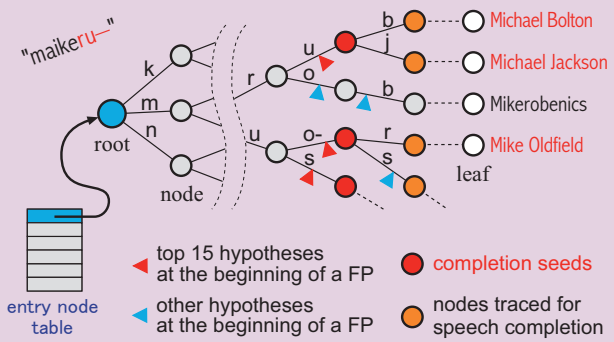
Speech Recognizer

❑ Provide a list of completion candidates from an uttered fragment

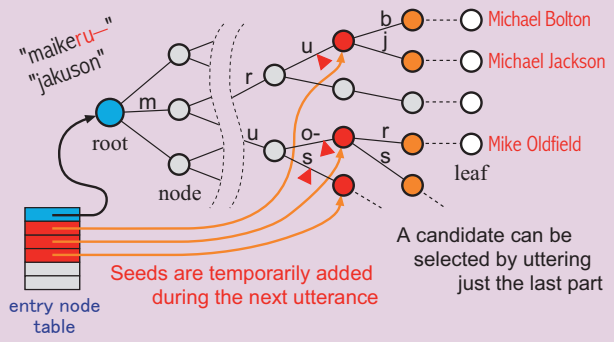
- Extend HMM-based speech recognizer *"niNja"*
- Send the results to the Interface Manager
  - Recognition results: At the utterance end
  - Completion candidates: During the utterance

❑ Generate candidates when FP is detected

- Trace from the completion seeds to the leaves



- Recognize last-part fragments



Experimental Results

❑ Tested with 45 subjects (24 male / 21 female)

- System vocabulary: 521 entries
  - Names of 179 Japanese musicians and 342 of their songs

❑ Evaluate whether the subjects preferred to use speech completion

- When a subject input a set of name entries written on a paper sheet
  - ➡ Average usage frequency: 74.2%
- When a subject had to recall and input vaguely remembered entries
  - ➡ Average usage frequency: 80.4%

❑ Subjective questionnaire results

- The assistance of candidate listing was helpful and easy to use
- The speech completion made it easy to recall and input uncertain phrases
- 80% of the subjects wanted to use the speech completion in the future

Masataka Goto, Katunobu Itou, and Satoru Hayamizu

National Institute of Advanced Industrial Sci. and Tech. (AIST)

Video clips:
http://staff.aist.go.jp/m.goto/ICSLP2002/

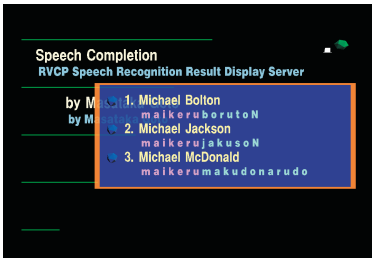
Snapshots

Foreign names are written or pronounced in the Japanese style

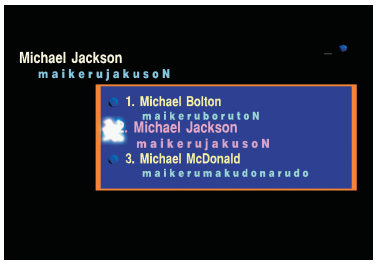
Forward Speech Completion

(1) Uttering "maikeru—"

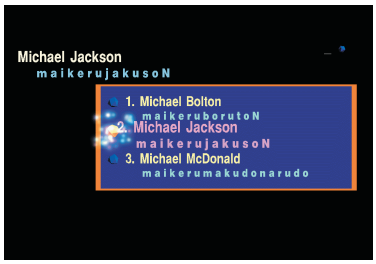
(2) During a filled pause "ru—"



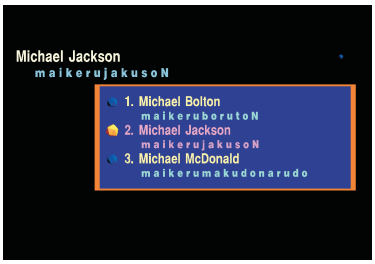
(3) A pop-up window containing completion candidates appears



(4) Uttering "No. 2"

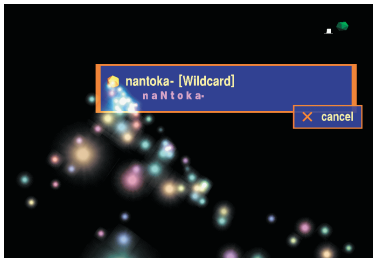


(5) The second candidate is highlighted and bounces

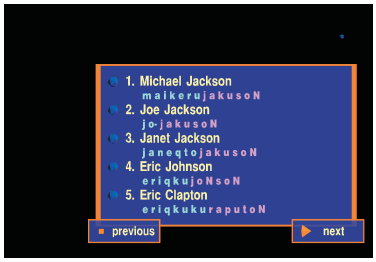


(6) The selected candidate is determined as the recognition result

Backward Speech Completion



(a) After uttering a *wildcard keyword* "nantoka—," a pop-up window appears



(b) After uttering "jakuson," a candidate window appears

(c) After uttering "No. 1," the first candidate is determined as the result

Summary

❑ Propose a new speech interface function "Speech Completion"

- Make use of nonverbal speech info. (filled pause)
- Filled pause is a good *"Speech TAB"* trigger
  - Can be detected independently of recognizer
  - Naturally used in human-human communication
- Intuitive enough to be used w/o any training
- Can be immediately applied to various speech applications

➡ Become as indispensable in speech IFs as text completion is in good text-based IFs

Future Directions

❑ Current speech input vs. keyboard input

- Speech recognizers have dealt with only a part of the normal letter keys

❑ Speech completion opens up new vistas

- Role of the special key  is triggered by the filled pause
- Assign other nonverbal information (ex. pitch, speech rate) to special keys

2002/09/19 ICSLP 2002 poster