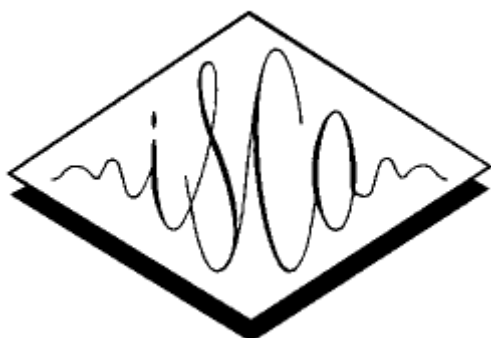


ISCA Archive

<http://www.isca-speech.org/archive>

**EUROSPEECH
2001
Scandinavia
7th European
Conference on
Speech
Communication
and Technology
2nd
INTER SPEECH
Event**



**Aalborg, Denmark
September 3-7, 2001**

Multi-Scale Retrieval in MEI: An English-Chinese Translingual Speech Retrieval System

Wai-Kit Lo (1), Patrick Schone (2), Helen M. Meng (1)

**(1) The Chinese University of Hong Kong, Hong Kong
(2) Department of Defense, USA**

This paper presents a multi-scale retrieval approach in MEI (Mandarin-English Information), an English-Chinese cross-lingual spoken document retrieval (CL-SDR) system. It accepts an entire English news story (from newspaper text) as the input query, and automatically retrieves "relevant" Mandarin news stories (from broadcast audio). This allows the user to search for personally relevant content across the language and media barriers - a cross-lingual and cross-media retrieval task. MEI advocates a multi-scale paradigm for the retrieval task. Multiscale refers to the use of both words and subwords (Chinese characters and syllables) for retrieval. Words offer lexical knowledge to enhance precision, and subwords can potentially alleviate some prevailing problems in CL-SDR, e.g. open vocabularies in translation and recognition, out-of-vocabulary words in audio indexing, and ambiguities in Chinese homophones and word tokenization. We present techniques for word-subword fusion, which improved retrieval performance in our experiments with the Topic Detection and Tracking collection.

[Full Paper](#)

Bibliographic reference. Lo, Wai-Kit / Schone, Patrick / Meng, Helen M. (2001): "Multi-scale retrieval in MEI: an English-Chinese translingual speech retrieval system", In *EUROSPEECH-2001*, 1303-1306.