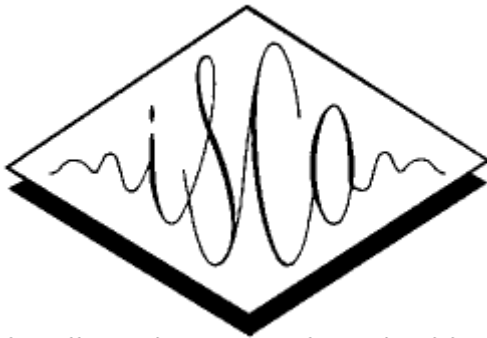**ISCA Archive**

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

EUROSPEECH 2003 - INTERSPEECH 2003
8th European Conference on Speech Communication and Technology

**Geneva, Switzerland
September 1-4, 2003**

Spoken Cross-Language Access to Image Collection via Captions

Hsin-Hsi Chen

National Taiwan University, Taiwan

This paper presents a framework of using Chinese speech to access images via English captions. The formulation and the structure mapping rules of Chinese and English named entities are extracted from an NICT foreign location name corpus. For a named location, name part and keyword part are usually transliterated and translated, respectively. Keyword spotting identifies the keyword from speech queries and narrows down the search space of image collections. A scoring function is proposed to compute the similarity between speech query and annotated captions in terms of International Phonetic Alphabets. The experimental results show that the average rank and the mean reciprocal rank are 2.04 and 0.8322, respectively, which is very close to the best performance, i.e., 1, for both average rank and mean reciprocal rank.

[Full Paper]

Bibliographic reference. Chen, Hsin-Hsi (2003): "Spoken cross-language access to image collection via captions", In *EUROSPEECH-2003*, 2749-2752.