PodCastle and Songle:  
Crowdsourcing-Based Web Services for  
Spoken Content Retrieval and Active Music Listening  

Masataka Goto  
Hiromasa Fujihara  
Kazuyoshi Yoshii  
Jun Ogata  
Matthias Mauch  
Tomoyasu Nakano  
National Institute of Advanced Industrial Science and Technology (AIST)  
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan  
m.goto [at] aist.go.jp

ABSTRACT

In this keynote talk, we describe two crowdsourcing-based web services, PodCastle (http://en.podcastle.jp for the English version and http://podcastle.jp for the Japanese version) and Songle (http://songle.jp). PodCastle and Songle collect voluntary contributions by anonymous users in order to improve the experiences of users listening to speech and music content available on the web. These services use automatic speech-recognition and music-understanding technologies to provide content analysis results, such as full-text speech transcriptions and music scene descriptions, that let users enjoy content-based multimedia retrieval and active browsing of speech and music signals without relying on metadata.

When automatic content analysis is used, however, errors are inevitable. PodCastle and Songle therefore provide an efficient error correction interface that lets users easily correct errors by selecting from a list of candidate alternatives. Through these corrections, users gain a real sense of contributing for their own benefit and that of others and can be further motivated to contribute by seeing corrections made by other users.

Our services promote the popularization and use of speech-recognition and music-understanding technologies by raising user awareness. Users can grasp the nature of these technologies just by seeing results obtained when the technologies applied to speech data and songs available on the web.

Categories and Subject Descriptors

H.3.5 [Online Information Services]: Web-based services; H.5.5 [Sound and Music Computing]: Signal analysis, synthesis, and processing; Systems

Keywords

Multimedia retrieval, web services, spoken content retrieval, active music listening, wisdom of crowds, crowdsourcing

1. INTRODUCTION

Our goal is to provide end users with public web services based on speech recognition, music understanding, signal processing, machine learning, and crowdsourcing so that they can experience the benefits of state-of-the-art research-level technologies. Since the amount of speech and music data available on the web is always increasing, there are growing needs for the retrieval of this data. Unlike text data, however, the speech and music data itself cannot be used as an index for information retrieval. Although metadata or social tags are often put on speech and music, annotations such as categories or topics tend to be broad and insufficient for useful content-based information retrieval [1]. Furthermore, even if users can find their favorite content, listening to it takes time. Content-based active browsing that allows random access to a desired part of the content and facilitates deeper understanding of the content is important for improving the experiences of users listening to speech and music. We therefore developed two web services for speech and music, PodCastle (Figure 1) and Songle (Figure 2).

2. PODCASTLE

PodCastle (http://en.podcastle.jp for the English version and http://podcastle.jp for the Japanese version) [3–7, 9, 10] is a spoken document retrieval service that uses automatic speech recognition (ASR) technologies to provide full-text searching of the speech data in podcasts, individual audio or movie files on the web, and the video clips on the video sharing services YouTube, Nico Nico Douga, and Ustream.tv. PodCastle enables users to find English and Japanese speech data including a search term, read full texts of their recognition results, and easily correct recognition errors by
to improving performance, and (3) the improved performance leads to a better user experience, which encourages further use of the service at step (1) of this spiral. This is a social correction framework, where users can improve the performance by sharing their correction results over a web service. The game-based approach of Human Computation or GWAPs (games with a purpose) [12] like the ESP Game [13] often lacks step (3) and depends on the feeling of fun. In this framework, users gain a real sense of contributing for their own benefit and that of others and can be further motivated to contribute by seeing corrections made by other users. In this way, we can use the wisdom of the crowd or crowdsourcing [11] to achieve a better user experience.

Acknowledgments: We thank Youhei Sawada, Shunichi Arai, Kouichirou Eto, and Ryutaro Kamitsu for their web service implementation of PodCastle, Utah Kawasaki for the web service implementation of Songle, and Minoru Saka for the web design of PodCastle and Songle. We also thank anonymous users of PodCastle and Songle for correcting errors. This work was supported in part by CREST, JST.

5. REFERENCES