The computational architecture of the human auditory cortex

Chapter 6

Knowledge valorization

Introduction

In our everyday life we are constantly exposed to sounds from disparate sources. Our brain is able to use these sounds to accomplish remarkable tasks, such as recognizing somebody's voice, understanding speech, or enjoying music. The formidable nature of these tasks becomes evident when one considers the input signals that the auditory system must deal with, namely pressure waves with complex temporal patterns. To make sense of the myriad of sounds around us, the auditory system must transform these complex patterns into representations that facilitate the readout of behaviorally relevant information. The work presented in this thesis has provided new insights on *how* complex, real-life sounds are represented by the human auditory cortex. The following paragraphs will discuss potential social and economical implications of the outcomes of this thesis.

Improvement of systems for automatic sound recognition

Technology has become an integral part of our daily life and constant effort is being made towards the development of systems for automatic sound recognition, which can be employed in several practical areas, including:

- Speaker identification, i.e. recognizing a particular speaker by the characteristics of his voice (*voice biometrics*). Speaker identification can be used e.g. by security systems to verify the identity of a speaker and represents a quicker and more secure alternative to authentication through passwords, personal data or standardized questions.
- Speech recognition, i.e. the automatic identification of spoken words and their translation into text. The main goal of speech recognition algorithms is to allow reliable human-machine interaction via vocal communication. Although we are far from having a machine that is able to hold conversations with humans, current speech technology can be a useful tool for a variety of applications, including information retrieval (e.g. automated search in large audio-archives), dictation (e.g. note-taking systems), home appliance control, voice commands for smartphones.
- Music transcription, i.e. the process of reconstructing the notation of a piece of music from its recording.
- Environment monitoring: automatic sound recognition can be used for monitoring environments. For example, a home-monitor system could signal an intrusion by identifying the sound of windows being smashed.

Despite their potential widespread application, current sound recognition systems are far from being perfect and their accuracy needs to be improved, especially in situations when interfering factors like e.g. background noise, sounds variability, or reverberation are present. The current effort for improvement is reflected in the attention that automatic sound recognition - and artificial intelligence in general - is receiving from software giants like e.g. Google and Facebook, which are striving to secure the world-leading experts in the field of deep learning, a branch of machine learning that promises to bring the mining of big data to a new level (Hinton et al., 2012). In March 2013, Google hired Geoffrey Hinton, a computer science professor at the University of Toronto and leading researcher in the deep learning field (University of Toronto, Media Room, 2013). Similarly, in January 2014, Google acquired DeepMind Technologies, an artificial intelligence start-up specialized in deep learning (Gibbs, 2014). These are just a few examples showing how active the hunting for machine learning talents is. However, powerful machine learning algorithms represent only one aspect of automatic sound recognition, namely the classification stage. In the processing pipeline of sound recognition systems, classification is preceded by feature extraction (Cowling & Sitte, 2003), which represents an equally critical factor for the achievement of good recognition performance. The goal of feature extraction is to transform incoming sounds into a representation that highlights acoustic features relevant for classification. In other words, the feature extraction stage projects the input waveform into a feature space that is better suited for distinguishing between different sound classes. Numerous features spaces have been proposed and tested (e.g. Cowling & Sitte, 2003; Kleinschmidt, 2002). However, despite the many decades of research, automatic sound recognition systems still lack robustness in adverse acoustic conditions. Human listeners, on the other hand, can effortlessly recognize sounds even in the presence of noise and despite the acoustic variability inherent in real-life sound sources. If the goal is to match human listeners performance, then building algorithms that mimic the human brain is likely to be the most effective strategy. In this framework, the work presented in this thesis can have a significant role in reducing the man-machine gap in sound recognition performance. Our studies have shed light on the feature space that the human auditory cortex uses to represent real-life sounds. This knowledge can readily be incorporated into state-of-the-art automatic sound recognition systems to improve their current performance.

Translating our results into a brain-inspired software for automatic sound recognition requires several steps. The first step should be the integration of our brain-based feature extraction software into the framework of current sound recognition systems. This could be very much favored by collaborations with leading research groups in the field of automatic sound recognition. Commercialization opportunities could be generated by establishing contacts with companies potentially interested in our product. Among others, these may include Soundhound, Audio Analytic, Google, Facebook, Apple, Nuance Communications, Sejona R&D, Microsoft, IBM.

References

Cowling, M., & Sitte, R. (2003). Comparison of techniques for environmental sound recognition. *Pattern Recognition Letters*, *24*(15), 2895–2907. doi:10.1016/S0167-8655(03)00147-8

Gibbs, S. (2014, January 27). Google buys UK artificial intelligence startup Deepmind for £400m. *The Guardian.* Retrieved from http://www.theguardian.com/technology/2014/jan/27/google-acquires-uk-artificial-intelligence-startup-deepmind

Hinton, G., Deng, L., Yu, D., Mohamed, A., Jaitly, N., Senior, A., … Kingsbury, B. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition. *IEEE Signal Processing Magazine*, *29*(6), 82–97.

Kleinschmidt, M. (2002). Robust speech recognition based on spectro-temporal processing. *PhD Thesis,University of Oldenburg*.

University of Toronto, Media Room. (2013). U of T neural networks start-up acquired by Google [Press release]. Retrieved from http://media.utoronto.ca/media-releases/u-of-t-neural-networks-start-up-acquired-by-google/