ELSEVIER

Speech Communication 34 (2001) 1–2

SPEECH COMMUNICATION

www.elsevier.nl/locate/specom

Editorial

Noise robust ASR

This special issue of *Speech Communication* focuses on noise-robust automatic speech recognition. Most of the papers included in this special issue are based on contributions presented at the Workshop on Robust Methods for Speech Recognition in Adverse Conditions held in Tampere, Finland, in May 1999. The scientific committee of the workshop first selected a set of authors to be invited to prepare extended and revised versions of their workshop papers. These extended paper versions were then peer reviewed by well-known researchers in the field. Based on the review reports, 14 papers were eventually chosen to be published in this special issue.

The objective of this workshop was to bring together researchers from both universities and industrial research groups who approach the problem of noise-robust speech recognition from different angles. Over 140 participants from Europe, America and Asia attended this workshop. The high number of participants clearly indicates that the workshop topic was very up-to-date. Many researchers share the same opinion that much work still needs to be done in order to improve the performance of the current state-of-the-art speech recognition systems in adverse conditions. It is thus obvious that noise-robust speech recognition will also remain one of the most active research fields in speech recognition in the future.

Despite the steady progress made in the area of speech recognition and a high number of practical applications, it is widely acknowledged that recognition technology today is not at the desired level. A small survey was conducted among the workshop participants, and, when we asked people to name some major problems which have been solved in speech recognition, over 30% of the respondents stated that there are no fully solved problems in this field. When asked to name the major unsolved problems in speech recognition, over 30% of the respondents again listed noise robustness as the first item.

In order to reduce the performance gap between laboratory and real-world systems, we must take a critical look at the existing approaches to feature extraction, acoustic modeling, etc. and develop new unorthodox solutions which could eventually boost recognition performance up to an acceptable level in all operating conditions. The results obtained in the opinion survey also suggest that there is a need for new ideas. The fact that cepstral coefficients and continuous density HMMs are still the dominant technology was considered by almost 25% of the respondents to be one of the major disappointments in the short history of modern speech recognition. We also asked the participants about the maturity of different recognition tasks. On a "maturity" scale of 1–5 (5 being the most mature), the following average scores were obtained:

- speaker-dependent isolated word recognition – 4.3;
- large-vocabulary continuous speech recognition with speaker adaptation – 2.7;
- speaker-independent connected-digit recognition of unknown length digit sequences – 3.1;
- automatic recognition of spontaneous speech – 1.4.

These scores also support the general understanding that we do not have any fully solved problems in speech recognition, and more research work is required to improve performance even in the simplest recognition tasks.

The papers selected for this special issue cover all major themes of the Tampere workshop which also reflect the most important research directions in noise-robust speech recognition. These 14 papers cover the entire speech recognition chain from

PII: S0167-6393(00)00041-8

signal acquisition to system dialogue. Based on these special issue contributions and the papers presented in the workshop, we can try to review the status of noise-robust speech recognition today. In general, there appears to be no single research area which would dominate noise-robust speech recognition research, but performance improvement in realistic operating conditions is being attempted in various parts of speech recognition systems. As researchers are also continuously reporting improved results, one could easily conclude that everything is as it should be. However, since the reported error rate reductions are still fairly moderate and there is only a handful recognition tasks where the recognition accuracy or the user success rate is close to 100%, we should carefully consider what kind of actions are needed for making significant progress in the field. Despite the large diversity in past and current research topics, it is still unclear in which direction we should proceed for making a breakthrough in noise robustness.

There are many obstacles today which complicate noise-robust speech recognition research. Complex recognition systems make it increasingly difficult to combine different successful techniques. A large variety of recognition tasks and the lack of common evaluation databases make it next to impossible to compare different noise-robust approaches. It is apparent that the current situation is not optimal for developing noise-robust processing techniques which would universally provide improved results across different recognition tasks and databases. The harmonization of test databases and tasks is apparently one important future target so that comparing different systems could be alleviated. Recognition tasks should also be chosen such that high performance, up to a 100% recognition accuracy, can be expected to be achieved in different noise conditions. Researchers are far too often satisfied if the error rate is halved, although recognition accuracy is still well beyond the acceptable level. If we cannot solve noise robustness issues in simple recognition tasks first, it is unrealistic to assume that it could be done in more complicated systems.

I hope that this special issue gives a good overall description of the current research activities in noise-robust speech recognition. As Guest Editor, I want to thank all the authors for their efforts in extending and revising their workshop papers for the special issue. I am also indebted to the members of the scientific committee and many other volunteers for their help in reviewing all the manuscripts. Thanks are also due to the Editor-in-Chief, Hervé Bourlard, for his advice and assistance during the publication process of the special issue. Finally, I would like to thank Nokia and COST249 for their financial support, which made it possible to arrange the workshop in Tampere.

O. Viikki
Nokia Research Center, Speech and Audio Systems Laboratory, P.O. Box 100, 33721 Tampere, Finland
E-mail address: olli.viikki@nokia.com