Title: **Future work on speech recognition improvements**
Source: **Vodafone**
Agenda Item: **7.4 (SA 4: Speech Enabled Services)**
Presented for: **Discussion & Decision**

1.   Introduction

For the past year or so, SA 4 has been actively examining the relative performance of codecs with regard to Speech Recognition. Given that:

- large quantities of voice traffic are carried by 3GPP networks;
- a substantial quantity of voice traffic is originated from vehicles; and
- there are increasing amounts of legislation restricting the in-car use of mobile phones;

then it is sensible (and indeed necessary) to study techniques by which speech recognition performance can be improved in the CS domain. However, the SA 4 work has NOT carried out this critical task.

Instead, the SA 4 work appears to have been directed onto "non-real time Speech Recognition over GPRS". Why?

> Note: The only reason for this appears to be non-3GPP work (probably within ETSI project AURORA), which seemed to have believed that it was "too difficult to do anything with the CS domain".

Very recently, the results of the SA 4 "speech recognition" codec competition have become available. Although Vodafone has not had time to fully analyse the results, it is clear that not all companies are agreed on the outcome of the competition.

More worryingly, the arguments remain focused on "speech recognition over GPRS" and do not relate to the real world… where the commercial requirements are clearly associated with improving speech recognition in the CS domain.

2.   Commercial Requirements

The commercially useful use cases for "speech recognition over GPRS" appear to be very limited.

One original use case was for "voice activated dialling". This is where a file containing the speech clip was sent to a PS domain network server which retrieved the phone number and sent it back as data to the mobile. The business rational behind this use case was that CS domain "operator

connected calls" always had problems with charging the correct tariff. However, early CAMEL functionality removes this tariffing restriction!

Other PS domain uses cases seem to be much more easily satisfied by CS domain voice calls interacting with an Automatic Speech Recognition Server that generates SMSs, MMSs, or WAP traffic towards the terminal. These use cases include Voice Activated Dialling; reading out SMSs/emails; downloading selected emails to the mobile; reading out or downloading weather forecasts; retrieving computer files; etc.

Overall, the vast majority of commercial use cases relate to the necessity to provide decent quality speech recognition in the CS domain. However the work in SA 4 has not been targeted at this requirement. Why?

> Note: it is clearly beyond the remit of the stage 1 description to mandate the CS, PS or IMS domain.

3. Technical Feasibility of CS domain solutions

The TR 23.877 from SA 2 indicates a couple of techniques for controlling speech recognition enhancements in the CS domain.

> Note: while commercial deployments of TFO may currently be limited, the current TR 23.977 on Bandwidth and Resource Savings indicates that (at least for operators with GSM spectrum) TFO is a fundamental component of TrFO, and, obviously, TFO is a fundamental component for AMR-WB. Hence TFO is likely to be available within the timescales of any new work codec development work within 3GPP.

But, mechanisms for improving speech recognition performance in the CS domain have not been studied by SA 4. Vodafone believe that this situation should be corrected.

4. Proposal

In line with existing agreements that 3GPP work should be commercially focussed, it is proposed that SA task SA 4 with the role of evaluating the costs and benefits of mechanisms for improving speech recognition in the CS domain.

Information to use within this study includes (but is not limited to) the results of the DSR vs AMR speech recognition codec competition, and, the TR 23.877.