Technical Specification Group Services and System Aspects **TSGS#22(03)0663**

Meeting #22, Maui, Hawaii, USA, 15-18 December 2003

Presentation of Technical Report to TSG

Presentation to:	**TSG SA Meeting #22**
Document for presentation:	**TR 23.877 "Speech Enabled Services" Version 1.0.0**
Presented for:	**Information**

Abstract of document:

TR 23.877v1.0.0 addresses architectural aspects of speech-enabled-services (SES). It is presented to SA#22 plenary for information because SA 2 considers it to be more than 50% complete.

The TR identifies a number of factors which might improve speech recognition performance and then discusses architectural and signalling mechanisms to implement them.

Changes since last presentation to TSG SA:

This is the first time this TR is presented to TSG SA.

Outstanding Issues:

The following outstanding issues are due to be addressed in future SA2 meetings:

- Privacy
- Security
- IMS roaming and charging
- Analysis and Conclusion section
- The analysis on impacts of the improving speech recognition performance techniques will be completed.

Contentious Issues:

None

3GPP TR 23.877 V1.0.0 (2003-12)

Technical Report

3rd Generation Partnership Project;
Technical Specification Group Services and System Aspects
Architectural Aspects of Speech Enabled Services;
(Release 6)

Keywords

<keyword[, keyword]>

3GPP

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

http://www.3gpp.org

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

1 Scope

The purpose of this technical report is to analyse the service requirements for speech enabled services as defined in 3GPP TS 22.243 Speech recognition framework for automated voice services; Service aspects (Stage 1).

The service requirements on Speech Enabled Services defined in TS 22.243 may place architectural requirements on the CS and PS domains and the IMS. Additionally various techniques for improving speech recognition performance may be identified which have architectural impact.

The objective of TR is to document these architectural impacts.

2 References

The following documents contain provisions that, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document.*

[1] 3GPP TR 21.905: " Vocabulary for 3GPP Specifications ".

[2] 3GPP TS 22.060: " General Packet Radio Service (GPRS); Service description; Stage 1".

[3] 3GPP TS 22.243: "Speech recognition framework for automated voice services Service aspects (Stage 1) ".

[4] 3GPP TS 23.060: " General Packet Radio Service (GPRS); Service description; Stage 2 ".

[5] 3GPP TS 23.228: " IP Multimedia (IM) Subsystem - Stage 2 ".

[6] 3GPP TS 23.002: "Network Architecture".

[7] 3GPP TS 29.007: "General Requirements on interworking between the PLMN and the ISDN or PSTN".

[8] 3GPP TR 23.910: "Circuit Switched Data Bearer Services".

[9] 3GPP TR 45.009: “GSM/EDGE Radio Access Network; Link adaptation”

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in 3GPP TS 22.174 [3] and the following apply.

Speech Recognition Mode: In this mode, the mobile modifies its speech/audio processing functions in a manner that is optimised for person to machine communication. When the mobile is not in this mode, the mobile’s speech/audio processing functions are assumed to be optimised for person-to-person communication.

3.2 Abbreviations

In addition to the abbreviations given in the remainder of this clause others are listed in 3GPP TR 21.905 [1].

ASR	Automatic Speech Recognition
SRM	Speech Recognition Mode
TFO	Tandem free operation
TRAU	Transcoder & Rate Adaptation Unit
TC	Transcoder

4 Techniques for Improving Speech Recognition Performance

Automatic Speech Recognition (ASR) platforms tend to perform the following sequence of operations: echo cancellation, feature extraction and interpretation. The echo cancellation process is used to permit a person to send commands to the ASR platform while the ASR platform is still playing voice announcements. (Without echo cancellation, the ASR platform cannot distinguish between its own "voice" and the voice of the user!) The feature extraction process extracts phonemes from the input speech signal and transforms them into words using acoustic models. The speech recognition engine then performs a search of the uttered words in the grammar created by GSL (Grammar Specific Language). The interpretation process is used to extract the semantic interpretation from the word sequence.

The following sub-sections describe techniques that might improve speech recognition performance. The gains of these techniques are still unclear and need to be clarified, before these techniques are standardized further. Later sections in this TR describe how the existing 3GPP system might be enhanced to provide signalling to control these techniques.

4.1 Additional Noise Suppression in Terminal

There are a number of factors that affect ASR performance, such as quality of signal, grammar size, acoustic confusability, etc. Challenges for speech recognition include background noise, side speech, hands free kits, and loss of signal quality due to compression. A large amount of wireless usage occurs in noisy environments. Reducing the impact of background noise seems to be useful in improving ASR performance.

Hence, ASR performance might be improved if the ASR unit can ask the UE to add extra noise suppression in the terminal.

4.2 Disable DTX for Improving ASR Performance

In normal case, Discontinuous Transmission (DTX) is used by the terminal to reduce the uplink transmissions. Silence Descriptor Frames are sent at a low rate and the TRAU uses them to send uplink comfort noise for improving normal speech quality.

A consequence of this is that during a Talk Spurt, the speech coder encodes the background noise, while during a non-voiced period; a different encoder encodes the background noise. This may give problems to the ASR platform when it uses the non-voice periods to estimate the background noise.

If the ASR unit could command the mobile to disable DTX, then the voice coder would always be used to encode the background noise and this might permit the ASR unit to perform more accurate noise estimation.

The degree of improvement in recognition performance would have to justify the additional power drain and increased interference due to non-DTX operation.

4.3 Indication of use/non-use of DTX

The ASR unit might find it beneficial to know whether or not DTX is being used. Extensions to TFO signaling could carry this information from the TRAU to the ASR unit.

4.4 Indication of Codec Mode to ASR Platform

In noisy wireless environments, the type of Codec used for compressing the speech may have some influence on speech recognition performance. Hence it might be useful if we could provide an indication of the used codec to the ASR platform.

4.5 Command Needed to Provide High Data Rate Codec for Speech Recognition Mode

The ASR recognition performance may drop significantly for codec rates below 8-10kbps. For this reason it should be possible to signal that during ASR sessions capacity optimization should not be considered for example, half rate channels should not be selected.

In reality, the actual rate used will depend on the radio channel conditions and degree of error (channel) protection required. This will maximize the actual quality of the speech presented to the recognizer.

4.6 TFO Needed in ASR Platform for AMR Wide Band Mode

AMR-WB can be implemented in the terminal to improve speech quality for normal voice calls. In order to move the Wideband signal to the ASR platform, TFO is likely to be needed.

4.7 Regional Accents and ASR Platforms

People from different regions speak with different accents and these people may suffer worse ASR performance.
Potentially, the HLR (or CAMEL platforms) could provide “regional accent information” to the VMSC that is passed onto the ASR platform (e.g. as modified A or B party numbers) to permit the ASR units to use “grammar” specific to that regional accent.

5. Impacts of Speech Recognition Improvements on CS Domain

Most of the techniques listed in section 4 can be implemented with either out of band signalling or in band signalling.
When using out of band signalling, the ASR unit would probably signal towards the MSC using Q.931 which then needs to be interworked into ISUP. At the MSC, additions to 24.008 and the Iu/A interface signalling would be needed. Within the RAN changes to RR/RRC and other protocols might also be needed.

For in-band signalling, extensions to the TFO protocol could be used by the ASR platform to talk directly to the TRAU, which could then map the signalling across to the BTS and onto the mobile.

From a practical point of view, the out of band signalling mechanism is more complex (extension to a lot of nodes and protocols as mentioned above) and it can be expected to take a long time to deliver this mechanism. Hence, this section of the TR focuses on extensions to the TFO in-band signalling mechanisms.

5.1 High Level TFO Signalling Flows in GSM/UMTS Network

The following procedures show how the ASR platform can use TFO to signal with nodes in the PLMN and with the mobile. The high level signaling flows in GSM and UMTS are shown in Fig.1 and Fig.2 respectively. The common procedures are described after the two diagrams.

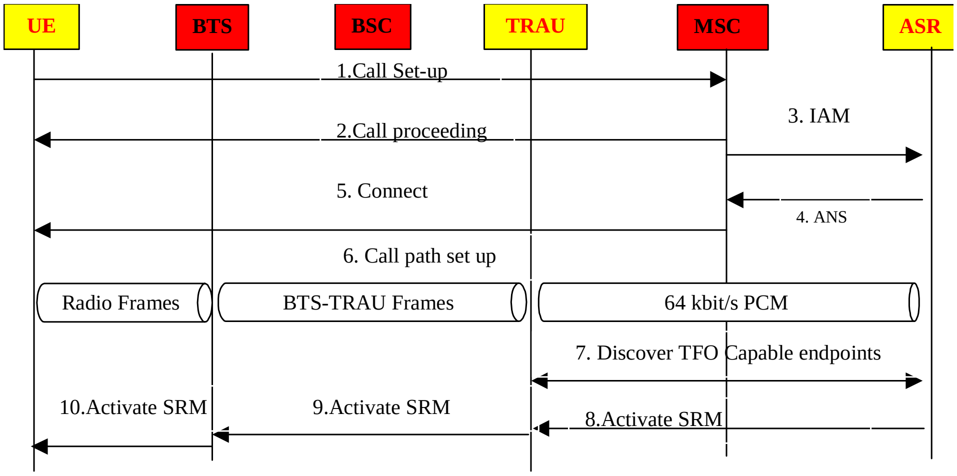

Fig.1 TFO Signalling Flows between GSM Terminal and ASR Unit

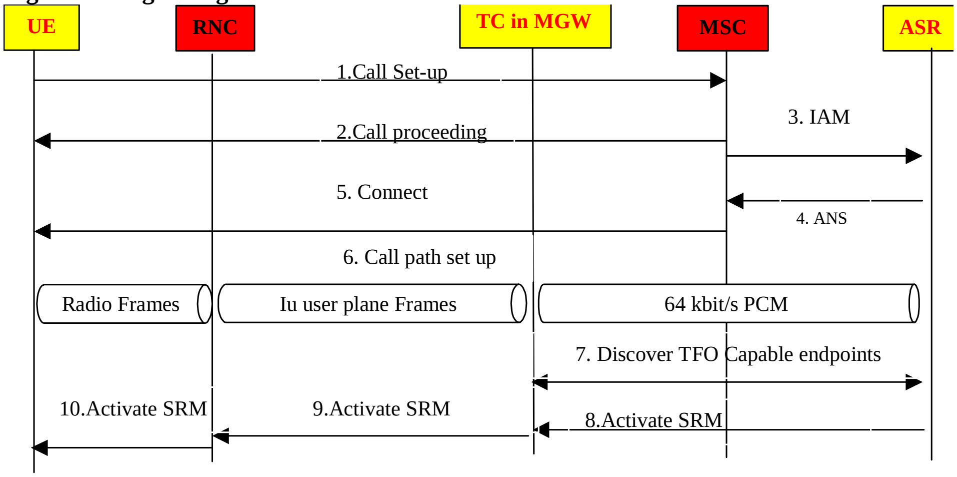

Fig.2 TFO Signalling Flows between UMTS Terminal and ASR Unit

The command signaling procedures in GSM/UMTS networks are described as follow:
1. Terminal sends call set-up message to MSC;

2. Terminal will receive call proceeding messaging;

3. MSC sends an IAM message towards ASR;

4. ASR sends the answer message back to MSC;

5. The connect message is sent from the MSC to the terminal;

6. Call path is established between terminal and ASR platform;

7. TFO capability is negotiated between TRAU (within BSC)/TC (within MGW) and ASR platform; at this point, the ASR unit could load a voice detection ‘library’ optimised for the codec indicated within the TFO signalling. Independently, the use of TFO provides the capability to pass a Wideband AMR voice stream to the ASR unit.

8. In the user plane, the ASR platform sends an instruction to Activate Speech Recognition Mode to the TRAU/TC;

9. TRAU/TC sends the Activate SRM instruction to the BTS/RNC;

10. BTS/RNC sends the Activate SRM instruction to the terminal. Within the terminal, this Activate SRM message can be used to perform several of the functions described in section 4, for example:
a) The mobile can add in extra noise suppression when it receives the Activate SRM message; and/or

b) DTX in terminal needs to be disabled for improving ASR performance.

5.2 Activation Mechanism for SRM Session

This section provides more details on how to provide the signalling for the ideas proposed in section 5.1. The following steps are used to activate the SRM session:

1. The SRU may pass the “activate SRM” instruction to the TRAU/TC by using extensions to the Tandem Free Operation signalling specified in 3GPP TS 28.062. The ASR regularly repeats the instruction to “activate SRM” [Refer to section 5.3].

2. The TRAU/TC passes the “activate SRM” instruction to the BTS/RNC by using extensions to the signalling specified in 3GPP TS 28.060/28.061 (for GSM) and 3GPP TS 25.415 (for UMTS).

3. For GSM, the BTS passes the “activate SRM” instruction to the mobile by, for example:
a) Using specific settings in the “frame stealing” bits, defining an unused combination of the stealing flags hu(B) and hl(B) to indicate that the burst contains speech and a command to activate SRM; or

b) Using particular codewords within the downlink SID frames, so that the command to activate SRM in conveyed by one or more SID frames sent during DTX periods; or

c) Using particular codewords within the downlink speech frames, so that the command to activate SRM in conveyed by frames containing speech; or

d) Using signalling messages sent on the FACCH from the BTS/RNC to the mobile (e.g. by adding information to a Physical Information message); or

e) Using signalling messages sent on the SACCH from the BTS/RNC to the mobile (e.g. by adding information to a System Information 6 message)

f) Use signalling within the RATSCCH channel (see 3GPP TS 45.009 [9]).
The UMTS mechanisms for passing the “activate SRM” instruction to the mobile are For Further Study.
Obviously, any stage 3 specification changes should be carefully designed to ensure that the above steps are fully ‘backwards compatibile’.

The details of the message and method of delivery is part of the stage 3 work in GERAN.

5.3 Deactivation Mechanism for SRM Session

There are several mechanisms by which the Speech Recognition Mode (SRM) could be deactivated in the mobile (and hence normal person-to-person communication reactivated). However, the one described here might be the simplest one, i.e. the ASR regularly repeats the instruction to “activate SRM”. Then if the mobile fails to receive the repeated ‘activate SRM’ instruction (e.g. N retransmissions missed or a timer expires), the mobile reverts to a normal person-to-person mode of speech processing.

If necessary, this technique could be extended so that the UE automatically reverts to person-to-person mode following a Handover Command.

5.4 ASR Platform Outside of the Mobile Network

Use of clean 64 kbit/s PCM links in the PSTN would mean that an ASR platform within an enterprise (e.g. Lufthansa customer care) could use TFO to interact with the TRAUs and mobiles within the PLMN.
This permits common development of ASR platforms towards all mobiles.

5.5 Charging

HPLMN should have the capability of supporting different charging mechanisms, such as subscription based, per-event basis, duration of voice tariff per second, content value/size, including variation of charging rate (e.g. free, standard rate, premium, etc). CAMEL should already provide the functionality for varying the charge rate.

All charging mechanisms will be required for pre-paid (CAMEL needed) and post-paid customers.

5.6 Roaming

Editor's Note: When roaming, should a codec other than that preferred by the UE not be supported by the roaming network then the default network codec is used.

5.6.1 Inbound Roaming

Inbound roaming means when a subscriber roams into a VPLMN. The VPLMN should have the capability of supporting inbound roaming subscribers, i.e. the subscriber should be able to use speech-enabled services in the visited network. This functionality can be provided using traditional short code - number translation features within the VMSC.

5.6.2 Outbound Roaming

HPLMN should have the capability of supporting out-bound roaming subscribers. When a subscriber registered to the speech enabled service roams abroad, the subscriber should be able to use his/her home speech enabled services, e.g. use of CAMEL to route calls to an ASR platform in the HPLMN.

6. Impacts of Speech Recognition Improvements on PS Domain

6.1 Charging - RADIUS Connected to ASR Platform

In the packet domain, charging mechanisms need to be provided by the ASR application. In order to generate a useful charging event record in the ASR platform, the ASR platform needs the permanent identity of the customer. ASR platform needs to be connected to RADIUS server and get a user's MSISDN from the IP address. Then ASR platform will be able to generate a CDR for charging against that MSISDN.

6.2 Fetching Regional Accent Information from MSISDN Associated Database

In section 6.1, RADIUS can provide a user's MSISDN information to the ASR platform. The MSISDN could be used to interrogate a database (e.g. HLR or an Administration platform) to obtain "regional accent" information.

6.3 Roaming

The use of a GGSN in the HPLMN permits out bound roamers to use their HPLMN services without modification.

7. Impacts of Speech Recognition Improvements on IMS Domain

Editor's Note: The decision on RTP/RTCP multiplexing or frame stealing about Codec information might impact on packet transmission over IMS.

The use of the DSR codec for speech recognition may have some impacts on IMS. Examples include:

a) Imagine the case where the ASR platform is running a voice activated dialling application. The mobile will need to use the DSR codec when talking to the ASR platform and switch to the AMR codec when being "through connected" to a human B party. How does the IMS signalling cause the codec to change? Two possibilities are discussed in the sections below.

b) When the mobile is using the DSR codec with the ASR platform, the AMR codec needs to be used in the downlink. Hence the SDP parameters included by the mobile need to reflect these uni-directional media flows. Note that this probably has an impact on many existing SIP implementations that expect to use the same codec in both directions.

The first of these issues also applies if the WB-AMR codec is used for speech recognition AND the B party only supports NB-AMR. The use of the NB-AMR codec for speech recognition appears to avoid this issue.

7.1 ASR Server Directly “Through Connects” the call to the B Party

An example of ASR platform acting as a SIP Back to Back User Agent is given in Fig.3.

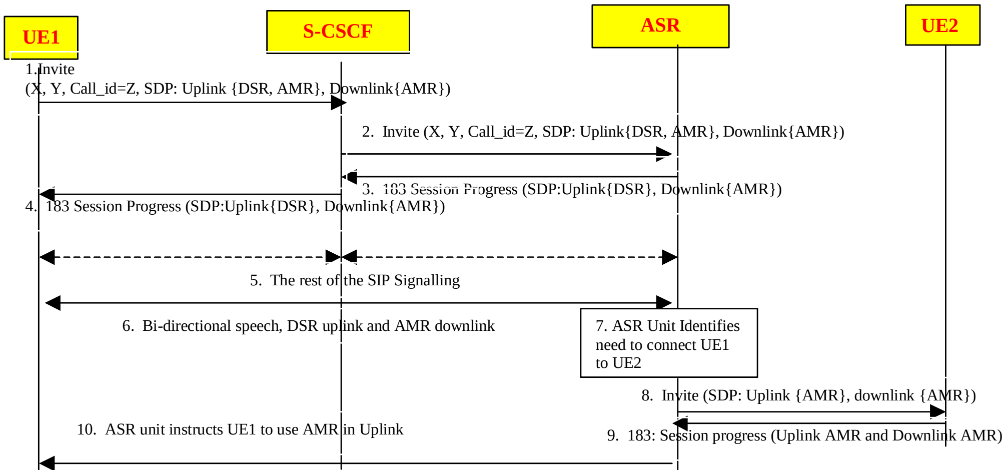
UE1
S-CSCF
ASR
UE2
1.Invite
(X, Y, Call_id=Z, SDP: Uplink {DSR, AMR}, Downlink{AMR})
2. Invite (X, Y, Call_id=Z, SDP: Uplink{DSR, AMR}, Downlink{AMR})
3. 183 Session Progress (SDP:Uplink{DSR}, Downlink{AMR})
4. 183 Session Progress (SDP:Uplink{DSR}, Downlink{AMR})
5. The rest of the SIP Signalling
6. Bi-directional speech, DSR uplink and AMR downlink
7. ASR Unit Identifies need to connect UE1 to UE2
8. Invite (SDP: Uplink {AMR}, downlink {AMR})
9. 183: Session progress (Uplink AMR and Downlink AMR)
10. ASR unit instructs UE1 to use AMR in Uplink

Fig.3 ASR Server Acting in “Through Connection Mode”

Step-to-step Procedures:
1. UE1 sends Invite to S-CSCF and SDP offers DSR and AMR in the uplink and AMR in the downlink;

2. S-CSCF sends Invite to ASR platform;

3. ASR platform sends 183 Session Progress back to S-CSCF; SDP indicates DSR in the uplink and AMR in the downlink;

4. S-CSCF sends 183 Session Progress back to UE; SDP indicates DSR in the uplink and AMR in the downlink;

5. The rest of the SIP signalling are used for setting-up a bi-directional speech path;

6. After further SIP signalling, a bi-directional speech path set up: DSR codec in the uplink, AMR codec in the downlink;

7. ASR unit identifies need to attempt to connect UE1 to UE 2;

8. ASR sends Invite to UE2 with initial SDP offer: SDP indicates AMR in the uplink and downlink;

9. UE2 sends Session progress with offered SDP (AMR uplink and AMR downlink) back from UE2

10. ASR sends instruction to UE1 and needs UE1 to swap to AMR. However, current SIP standards do not seem to permit this. Even if, in step 4, the ASR unit returns SDP parameter indicating that DSR should be used but that AMR is also permitted (in the uplink), then there does not seem to be any method for the ASR unit to command the switch from DSR to AMR. Further study is needed to resolve this issue if DSR and/or WB-AMR is used for speech recognition in the IMS domain.

7.2 ASR Server Acting in Call Transfer Mode

An example of ASR platform acting as in Call Transfer mode is given in Fig.4.

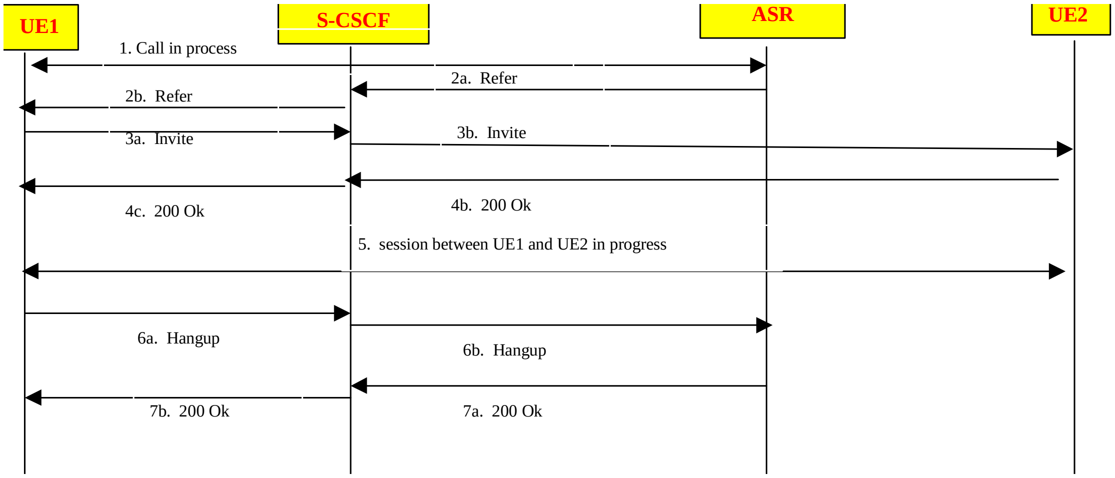

Fig. 4 ASR Server Acting in REFER Mode

Step-to-step Procedures:

1. Call in process between UE1 and ASR;

2. ASR sends “REFER call to UE 2” message to UE1;

3. UE1 sends Invite to UE2;

4. UE2 sends 200 Ok to UE1;

5. Session between UE1 and UE2 is in progress;

6. UE1 sends Hangup message to ASR unit;

7. ASR sends 200 Ok message back to UE1.

The problem with this mode is that it is prevents interaction with the ASR once the first session is completed. Conversely, the “through connect” model permits the ASR unit to reconnect itself when, say, the # key is pressed, or the user shouts a certain key word. (This would permit the customer to use the ASR to connect them to 30 seconds of weather forecast then use the ASR unit to connect them to CNN-news, then use the ASR unit to listen to voice mail).

7.3 Charging

7.4 Roaming

8. Security

9. Privacy

10. Analysis and Conclusions

Annex <A> (normative): <Normative annex title>

Annex <B> (informative): <Informative annex title>

Annex <X> (informative): Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2003–08					Skeleton created at SA2#34, Brussels, Belgium		0.0.0
2003-10					TR 23.877 created at SA2#35, Bangkok, Thailand		0.1.0
2003-11					TR 23.877 created at SA2#36, New York, USA		0.1.1
2003-11					TR 23.877 created at SA2#36, New York, USA		0.2.0
2003-12					First presentation for Information	0.2.0	1.0.0