

# Speechsc Agenda

- 0900 - Agenda Bashing/Charter Review (Chairs)
  - MINUTE TAKER!! BLUE SHEETS!
- 0910 - Work Roadmap & Timeline (Chairs)
  - Charter
  - Work Items
  - Timeline for Work Items
- 0930 - Discuss requirements document (draft-burger-speechsc-reqts-00.txt)
  - open issues
  - assess readiness for WGLC
- 1000 - Discuss usage cases document  
(draft-maes-speechsc-use-cases-00.txt)
- 1030 - Discuss web services document  
(draft-maes-speechsc-web-services-00.txt)
  - adopt as WG task?
- 1045 - Brainstorm on getting the protocol definition document written
  - authors/editors?
  - design baseline
  - pieces of draft-shanmugham-mrcp-01 (now expired) to adopt?
- 1115 - Wrap-up and next steps

# WG Status - Charter

- WG renamed from CATS to SPEECHSC by IESG
- Scope initially restricted to ASR, TTS and SV, expand later as we demonstrate we can deliver
- No mandate to use RTSP, but strong guidance to re-use/extend existing protocol rather than start from scratch
- Coordination with ETSI Aurora, ITU SG16, and W3C explicitly called out

# Work Items

- From Charter:
  - Jul 02 Informational Requirements RFC Published
  - Dec 02 Submit Internet Draft(s) Analyzing Existing Protocols
  - Dec 02 Submit Internet Draft Describing New Protocol (if required)
  - Mar 03 Drafts to IESG

# Milestone Timeline

- Would like to do WGLC on requirements by early next month
  - Discussion of requirements draft should give indication if this is do-able
- Would like to kick off work on protocol analysis immediately following WGLC of requirement
  - Discuss best way to accomplish this milestone

# Requirements Draft

## Open Issues (1)

- Should the TTS Server infer the text is SSML by detecting a legal SSML document, or must the protocol tell the TTS Server the document type?
  - Better to always require explicit content-type header (Burnett)
  - **Proposed Resolution:** require content type header
- Should we allow (or require) the TTS Server to use long-lived control channels?
  - Must at least allow (Burnett)
  - **Proposed resolution:** protocol **MUST** allow establishment of long-lived control channel, but not **REQUIRE** it always be up
- Should we allow for session parameters, like prosody and voicing, as is specified for MRCP over RTSP [7]?
  - Need this, and more, plus extensibility (Maes)
  - Need to be able to set these on per-session basis (Burnett)
  - **Proposed Resolution: yes.** Will craft appropriate wording

# Requirements Draft

## Open Issues (2)

- Should we allow for speech markers, as is specified for MRCP over RTSP
  - Need this, and needs to be efficient (Maes)
  - Isn't SSML adequate here (Burnett)
  - **Proposed resolution:** require ability to use SSML markers but leave it at that until SSML shown inadequate
- Should we allow the ASR Server to support alternative grammar formats? If so, we need mechanisms to specify what format the grammar is in, capability discovery, and handling unsupported grammars.
  - Yes, need this (Maes)
  - Don't need capability discovery (Burnett)
  - **Proposed resolution: yes**, at minimum have way to explicitly signal grammar format, but leave capability discovery out for now.
- Is there a need for all of the parameters specified for MRCP over RTSP [7]? Most of them are part of the W3C speech recognition grammar.
  - Yes, need to go beyond W3C grammar, plus also need extensibility (Maes, Burnett)
  - **Proposed resolution: Yes. MUST** be able to specify parameters on a per-session basis. Exact set to be decided as part of the protocol analysis and design phase. **SHOULD** also support parameter specification on a per-interaction basis

# Requirements Draft

## Issues raised on List (1)

- Requirements should cover other servers besides ASR, TTS, and SV/SR (Maes)
  - **Proposed resolution: no**, as current charter limits scope.
- Protocol must provide for serial and parallel composition of servers (Maes)
  - **Proposed Resolution: no:** Current charter limits scope, need to walk before we run, serial chaining runs afoul of OPES proxy issues Note as an area for investigation and possible enhancement in the future
- Does requirement to not reinvent RTSP, SIP/msuri restrict ability to use markers and other playout options like pacing? (Maes, Burnett)
  - **Proposed resolution:** will reword requirement to make it clear that e do not have to live within these confines to meet the other requirements
- Clarify what OPES requirement means – who exactly is operating on behalf of one user
  - **Proposed resolution:** Will clarify – intent is that it is client side of protocol – the IVR, PSTN gateway, etc.

# Requirements Draft

## Issues raised on List (2)

- Load balancing should be completely out of scope (Maes). Framework MUST NOT require external load balancing among engines (Burnett)
  - **Proposed resolution:** current words capture consensus for long discussion at BoF and on email list, but Burnett’s words seem useful and should be added
- Must be able to control language and prosody for plain text (Maes)
  - **Proposed resolution:** misunderstanding – get this with SSML, will clarify text
- Need “Full control” over TTS engine (Maes). VCR and other fine-grained control should be lower priority (Burnett)
  - **Proposed resolution:** ???
- Grammar specification must allow both by value or reference (Maes, Burnett)
  - **Proposed Resolution:** yes



# Requirements Draft

## Issues raised on List (3)

- Must handle prompting, recording, possibly utterance verification, utterance alignment, retraining, in addition for requirement to record for analysis (Maes)
  - Original requirement just for analysis/tuning of ASR, but extensibility to these other things desirable (Burnett)
  - Must not preclude capture no matter what (Burnett)
  - **Proposed Resolution:** design for extensibility, but no specific requirements on protocol other than recording for now.
- Grammar sharing
  - Mostly implementation issue (Maes)
  - two needs (1) ability to store large grammars that may have originally been sent by the client, and (2) an ability to reference grammars already known ("builtin") to the server (Burnett)
  - Also must not preclude grammar sharing across sessions (Burnett)
  - **Proposed resolution:** Adopt Burnett's description

# Requirements Draft

## Issues raised on List (4)

- Need to cover speaker enrollment, identification and classification, as well as speaker *recognition* as part of SV. Also multiple methods need to be accommodated - text-dependent, prompted, etc. (Maes, Burnett)
  - **Proposed resolution: yes.** Will add this to requirements
- Holding cross-utterance state is implementation option. Why a requirement here. Instead make a requirement on extensibility (Maes)
  - Protocol really need to explicitly accommodate this state holding both in the SRCP framework and in the application (Burnett)
  - **Proposed resolution:** requirement stands as is
- Need simultaneous ASR and SV (Burnett, Maes)
  - **Proposed resolution: yes.** Will add requirement to this effect but leave thew parallel decomposition as something we will look at but maybe leave out due to OPES issues.

# Plan of Action

- Reissue Requirements draft by end of next week
- Shoot for WGLC by early August, with agreed text to IESG before end of August
- **Meta Issue:** Do we fold in use cases to requirements document, issue it as a standalone RFC, or just use it to guide the protocol analysis and later design?