

vrSpeech

vrSpeech *type turn-id [speaker|utt_id] [confidence tone speech]*

Description

vrSpeech messages go together as a sequence with the same *turn-id*, representing the speech recognition, timing, and possible segmentation of a human utterance. vrSpeech *start* means the person has started to speak - indicated usually by pressing on a mouse button. vrSpeech *finished-speaking* means the person has stopped speaking for this turn, indicated usually by releasing the mouse button. vrSpeech *interp* indicates the text that the person said during this speech. vrSpeech *asr-complete* means that speech recognition and segmentation are finished for this turn, and other modules are free to start using the results without expecting more utterances for this turn. vrSpeech *partial* is an optional message, giving partial results during speech.

- vrSpeech *start turn_id speaker*
- vrSpeech *finished-speaking turn_id*
- vrSpeech *partial turn_id utt_id confidence tone speech*
- vrSpeech *interp turn_id utt_id confidence tone speech*
- vrSpeech *asr-complete turn_id*

Parameters

- *type*, indicates the type of vrSpeech message:
 - start
 - finished-speaking
 - partial
 - interp
 - asr-complete
- *turn_id* is a unique identifier for this whole turn. Must be the same across all messages in a sequence.
- *speaker* is the speaker of this message, like 'user' in the Brad example.
- *utt_id* is the unique identifier for this utterances. A turn can be segmented into multiple utterances. Historically this is always '1', and there is only one 'vrSpeech interp' message per turn. The current convention is that the speech recognizer / client will send out a message with 'utt_id 0', and then a possible segmenter will send out one or more messages from 1 to n, labelled in order. A segmenter is currently not part of the Toolkit. For 'vrSpeech partial' *utt_id* is an incremental interpretation of the part of the utterances that has been processed so far. It will likely have a value of p followed by an integer but any unique identifier should be ok.
- *confidence*, the confidence value for this interpretation. Should be between 0.0 and 1.0. Historically this is always 1.0.
- *tone*, the prosody - should be "high" "normal" or "low" - historically, this is always "normal".
- *speech*, the text of the (segment) interpretation. For 'utt_id 0' this will be the text of the full turn, for other segments this will be the text only for that segment.

Examples

This example indicates a speaker with the name 'user' starting to speak, saying 'hello how are you', and then stopping. Note that the speech is only analyzed / completed after the user has stopped speaking. Also note that all listed messages are needed to process a single utterance and notify the rest of the system of the results.

- vrSpeech *start test1 user*
- vrSpeech *finished-speaking test1 user*
- vrSpeech *interp test1 1 1.0 normal hello how are you*
- vrSpeech *asr-complete test1*

Sending Components

- [AcquireSpeech](#)

Receiving Components

- [NPCEditor](#)
- Natural Language Understanding modules
- Modules that want to know about speech timing information or spoken text

Related Messages