

Voice Commands to Control Recording Sessions

Handout 1: Background, Motivation and Design

Overdub Recording Workflow
Detailed description of Apparatus:
Audacity and “SayPlay”

John “Marty” Goddard

Introduction:

Goals

- Provide hands-free tape operation
- Allow musicians to record themselves

Historical

- “Put That There” gestures and voice 1981
- Voice Navigator for Macintosh Musicians 1989

5 Most Important Sources

#1 Lee, K.F. 1988 “Large-vocabulary speaker-independent continuous speech recognition- The SPHINX system”

#2 Lemon, O., Gruenstein, A., 2004. "Multithreaded context for robust conversational interfaces: Context-sensitive speech recognition and interpretation of corrective fragments.”

#3 Nakano, Teppei 2008 “Flexible Shortcuts: Designing a New Speech User Interface for Command Execution”

#4 Rudnicky, Alexander I. 1989 “*The Design of Voice-Driven Interfaces*”

#5 Yavelow, Peter 1989, “Voice Navigation for the Macintosh Musician”

Large-vocabulary speaker-independent continuous speech recognition: The SPHINX system

Lee, Kai-Fu, Ph.D.

Carnegie-Mellon University, 1988

Likely the basis for Microsoft Speech Recognition, as he was hired by Microsoft to work on Speech Recognition. Highly referenced source for successful Dictation speech recognition systems. The strategy was to implement the previously competing features of having both Large Vocabulary and Continuous (rather than keyword) speech recognition for dictation recognition systems. Speaker independence was also addressed, but with less success without some training, than given some amount of training.

Multithreaded Context for Robust Conversational Interfaces: Context-Sensitive Speech Recognition and Interpretation of Corrective Fragments

OLIVER LEMON
Edinburgh University

ALEXANDER GRUENSTEIN
Stanford University

Conversations are by nature multi-threaded, meaning that several lines of thought can be going on simultaneously. Context sensitivity limits the possible referents to within a context.

Corrective Fragments refer to the user making a correction to the current actions by issuing a revised command.

Flexible Shortcuts: Designing a New Speech User Interface for Command Execution

Teppei Nakano

Department of Computer Science,

Waseda University

3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan

teppei@pcl.cs.waseda.ac.jp

This research is focused on the problem of user's not always remembering the precise command phrase. Continuous Keyword Input allows the user to be more flexible in each spoken command. Results compare well against traditional Command and Control (C&C). Experiment compares control of Windows Media Player and Voice Chat (Skype), and is measured using subjective assessment.

The design of voice-driven interfaces

Alexander I. Rudnicky

**School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213**

Design the Language of the Command Set based on observations of users performing tasks by voice in unconstrained situations. Design facilities to “promote fluent interaction, error repair, and capability to introduce new (task-specific) words”. Voice controlled Spreadsheet application.

Voice Navigation for the Macintosh Musician

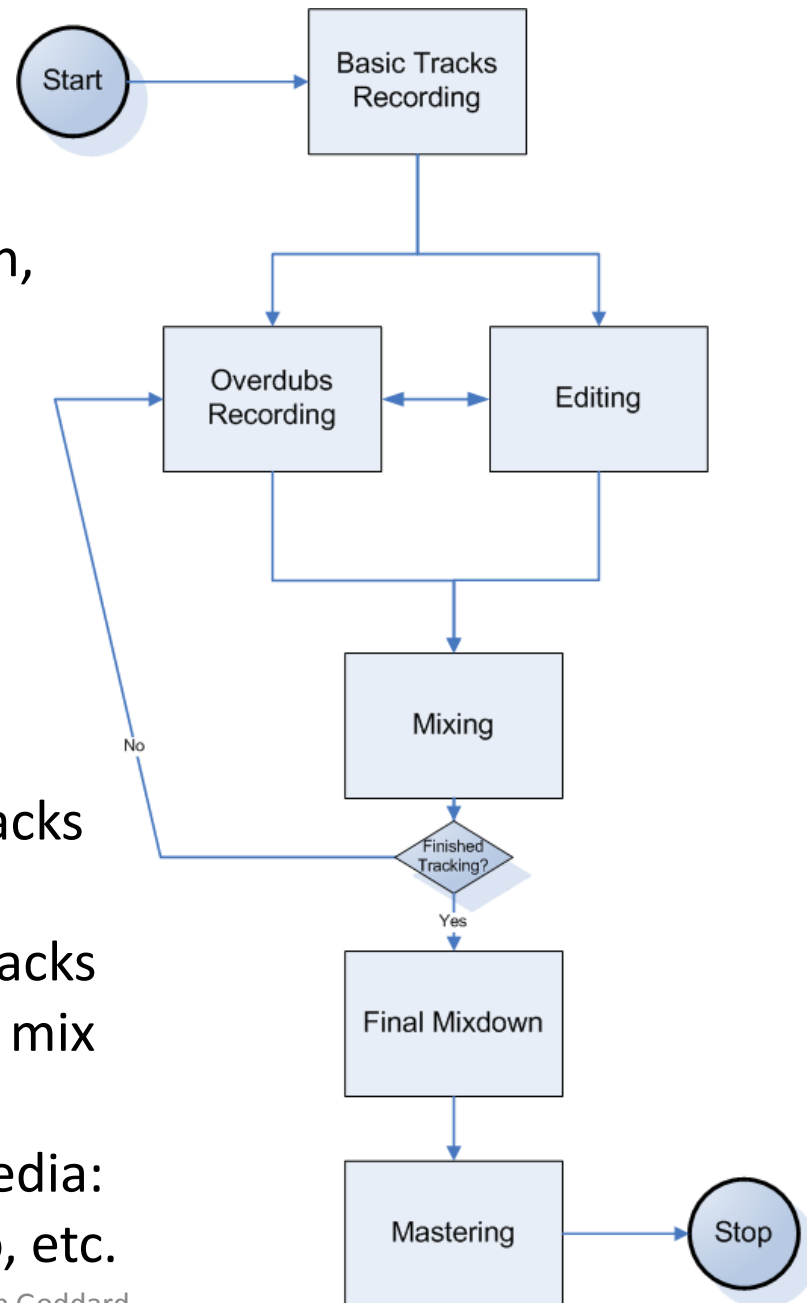
By Christopher Yavelow

Voice commands to aid in music production, calling up patches by name, as well as remote control. Emphasis is placed on having hands free to control music mixing.

No actual review of functioning software is published

Recording workflow

- Basic Tracks: Record initial Run-Through, one track per instrument/voice
- Overdubs: Add instrument or voice to Basic Track recording
- Editing: Cut, paste, shift, replace
- Mixing: Blend and balance recorded tracks
- Final Mixdown: After all recording of tracks is finished, make a final mix
- Mastering: adjust final mix to target media: CD, DVD, iTunes, MP3, radio, etc.



Overdub Recording Workflow:

(adding instruments and voices to existing recording)

Record Take: Begin Recording, perform music, stop recording

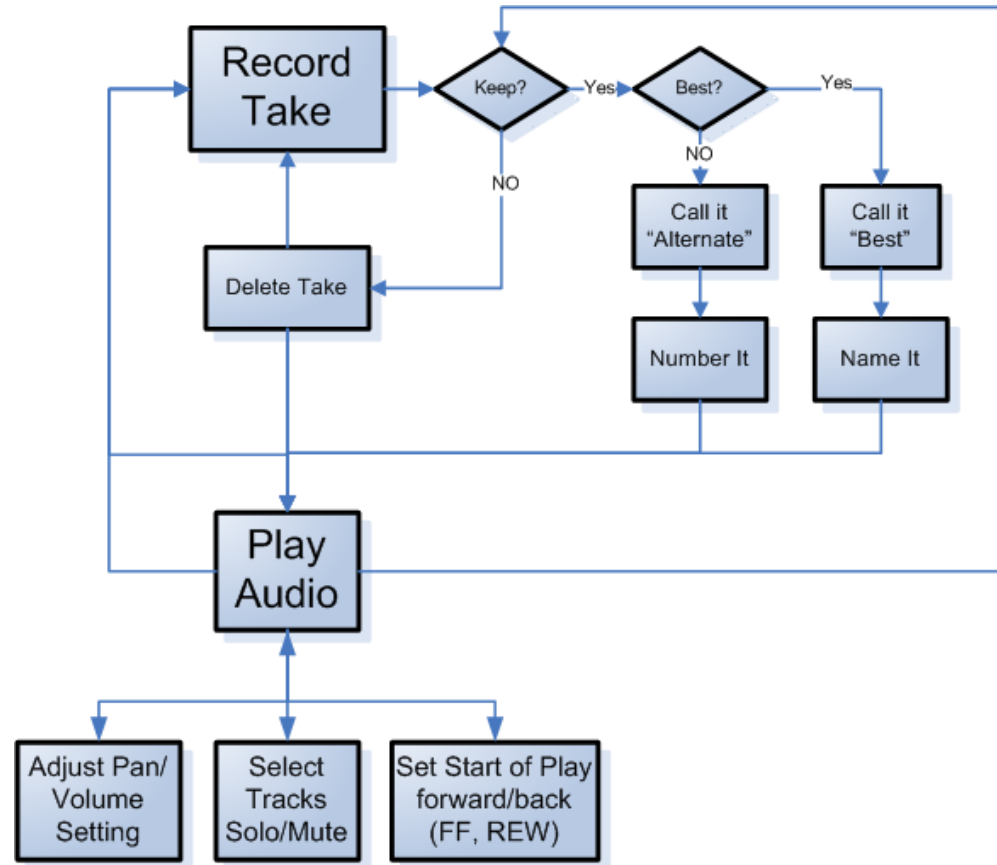
Delete take if not salvageable

Play Audio: Audition, or playback recorded tracks and listen for performance problems.

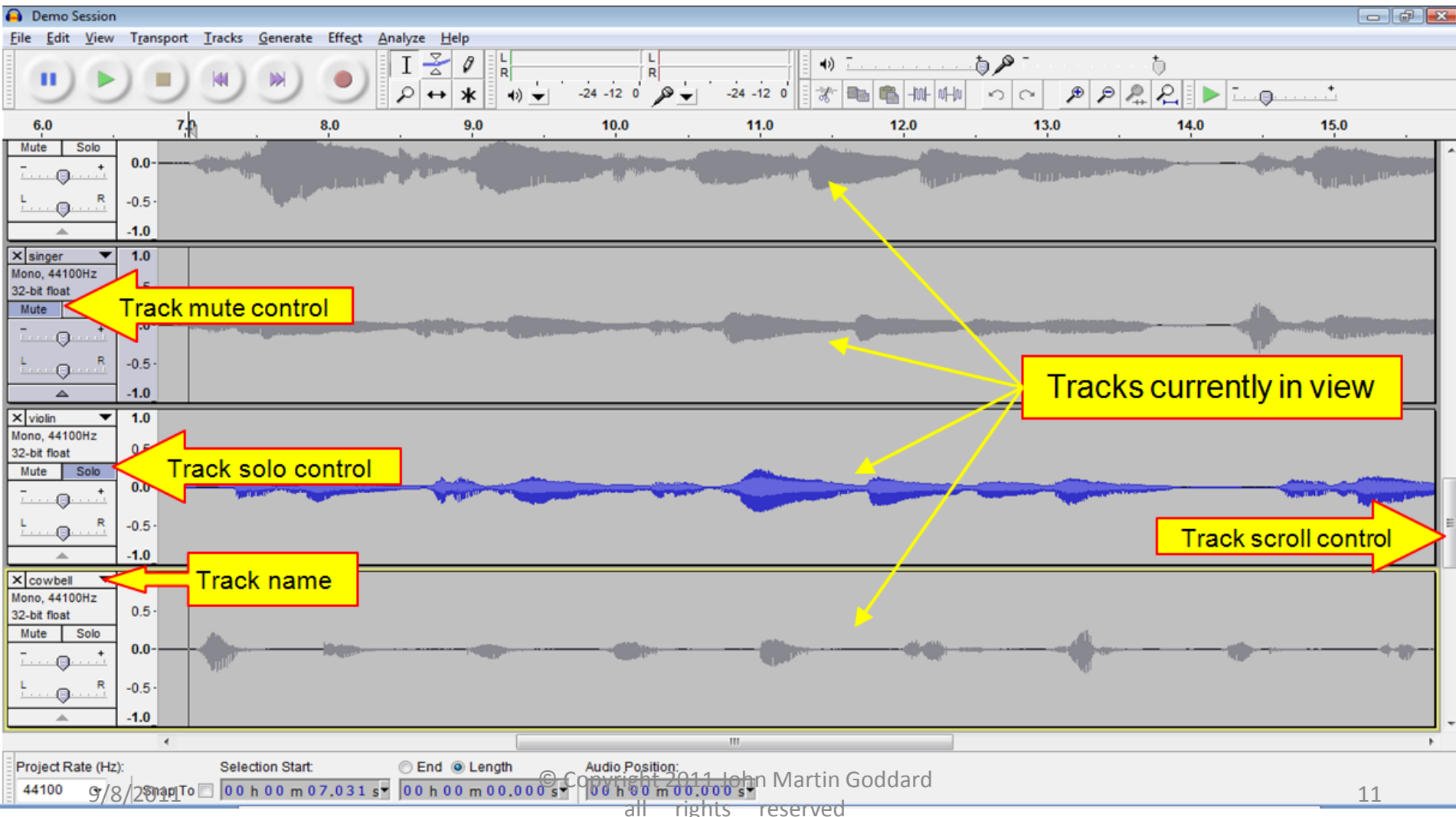
Solo: Selects track to hear, others are temporarily turned off

Mute: Temporarily turn off this track

Pan: shift left or right in stereo mix

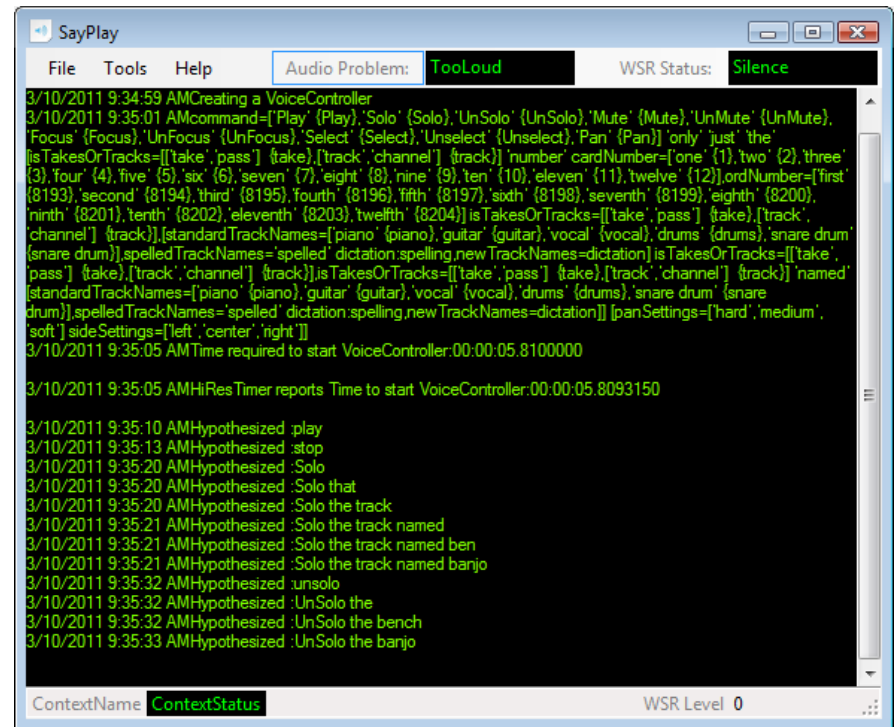


Audacity audio recorder/editor

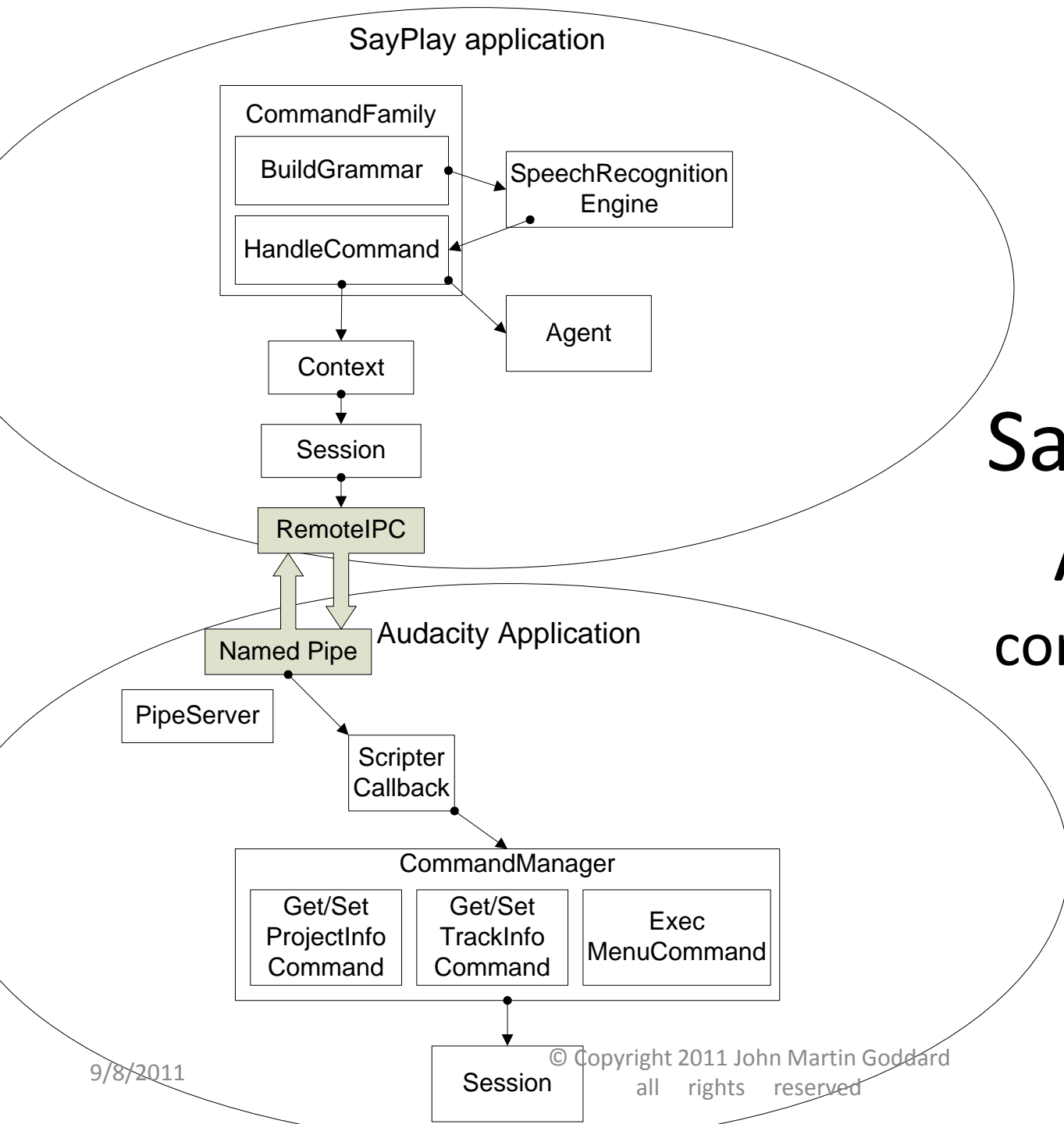


SayPlay: handles voice command events, sends formatted commands to Audacity

- Accepts Voice input
- Interacts with Speech Recognition Engine
 - Creates Grammar structures
 - Handles recognition events
- Sends commands to Audacity via Script Interface
- Logs command events and results to a text file



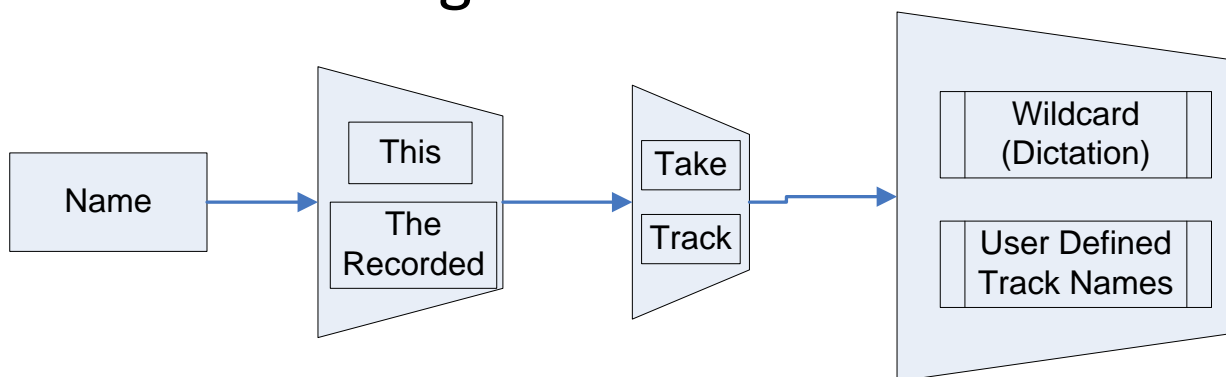
```
3/10/2011 9:34:59 AM Creating a VoiceController
3/10/2011 9:35:01 AM command=[Play] {Play}, {Solo} {Solo}, {UnSolo} {UnSolo}, {Mute} {Mute}, {UnMute} {UnMute},
{Focus} {Focus}, {UnFocus} {UnFocus}, {Select} {Select}, {Unselect} {Unselect}, {Pan} {Pan} only just the
{isTakesOrTracks=[[take, pass] {take}, {track, channel} {track}] number cardNumber=[one {1}, two {2}, three
{3}, four {4}, five {5}, six {6}, seven {7}, eight {8}, nine {9}, ten {10}, eleven {11}, twelve {12}], ordNumber=[first
{8193}, second {8194}, third {8195}, fourth {8196}, fifth {8197}, sixth {8198}, seventh {8199}, eighth {8200},
ninth {8201}, tenth {8202}, eleventh {8203}, twelfth {8204}] isTakesOrTracks=[[take, pass] {take}, {track,
channel} {track}], standardTrackNames=[piano {piano}, guitar {guitar}, vocal {vocal}, drums {drums}, snare drum
{snare drum}], spelledTrackNames=spelled dictation.spelling.new TrackNames=dictation] isTakesOrTracks=[[take,
pass] {take}, {track, channel} {track}], isTakesOrTracks=[[take, pass] {take}, {track, channel} {track}] named
[standardTrackNames=[piano {piano}, guitar {guitar}, vocal {vocal}, drums {drums}, snare drum {snare
drum}], spelledTrackNames=spelled dictation.spelling.new TrackNames=dictation]] [panSettings=[hard, medium,
soft] sideSettings=[left, center, right]]
3/10/2011 9:35:05 AM Time required to start VoiceController:00:00:05.8100000
3/10/2011 9:35:05 AM HiResTimer reports Time to start VoiceController:00:00:05.8093150
3/10/2011 9:35:10 AM Hypothesized :play
3/10/2011 9:35:13 AM Hypothesized :stop
3/10/2011 9:35:20 AM Hypothesized :Solo
3/10/2011 9:35:20 AM Hypothesized :Solo that
3/10/2011 9:35:20 AM Hypothesized :Solo the track
3/10/2011 9:35:21 AM Hypothesized :Solo the track named
3/10/2011 9:35:21 AM Hypothesized :Solo the track named ben
3/10/2011 9:35:21 AM Hypothesized :Solo the track named banjo
3/10/2011 9:35:32 AM Hypothesized :unsolo
3/10/2011 9:35:32 AM Hypothesized :UnSolo the
3/10/2011 9:35:32 AM Hypothesized :UnSolo the bench
3/10/2011 9:35:33 AM Hypothesized :UnSolo the banjo
ContextName ContextStatus WSR Level 0
```



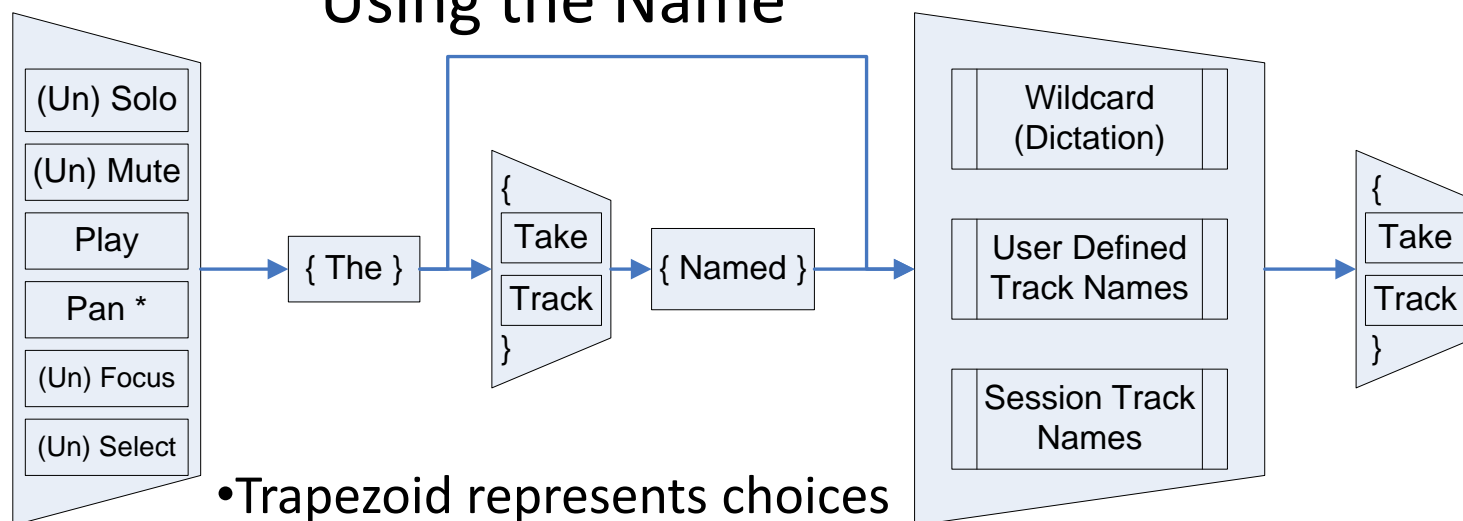
SayPlay and Audacity context diagram

Grammar Structures for Naming Tracks

Creating the Name



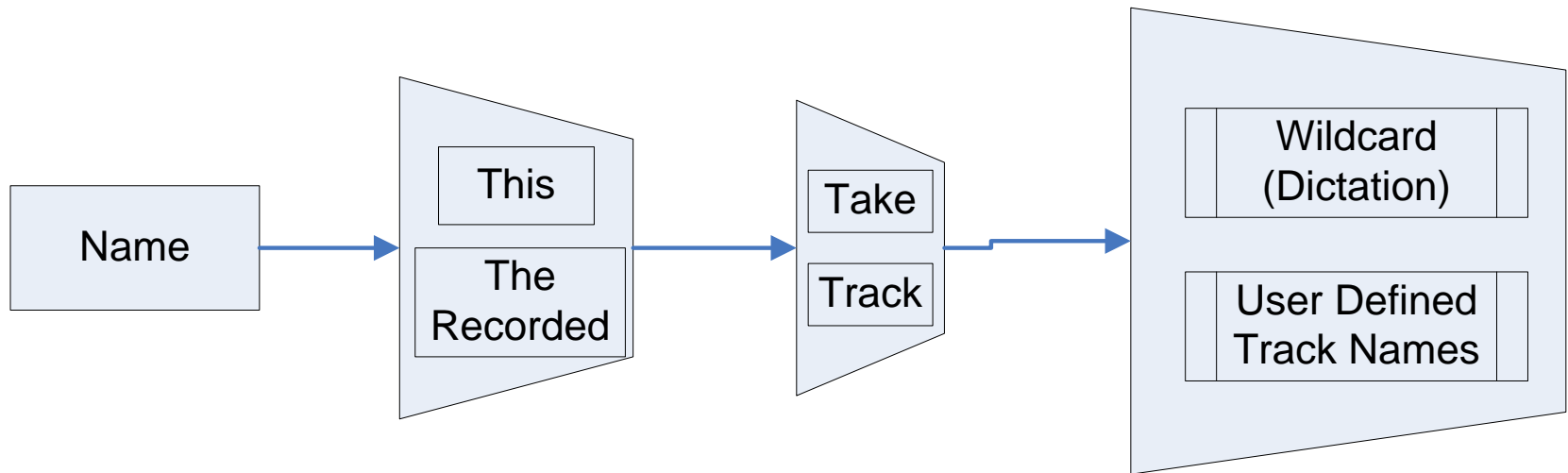
Using the Name



- Trapezoid represents choices
- { } Indicates optional word(s)

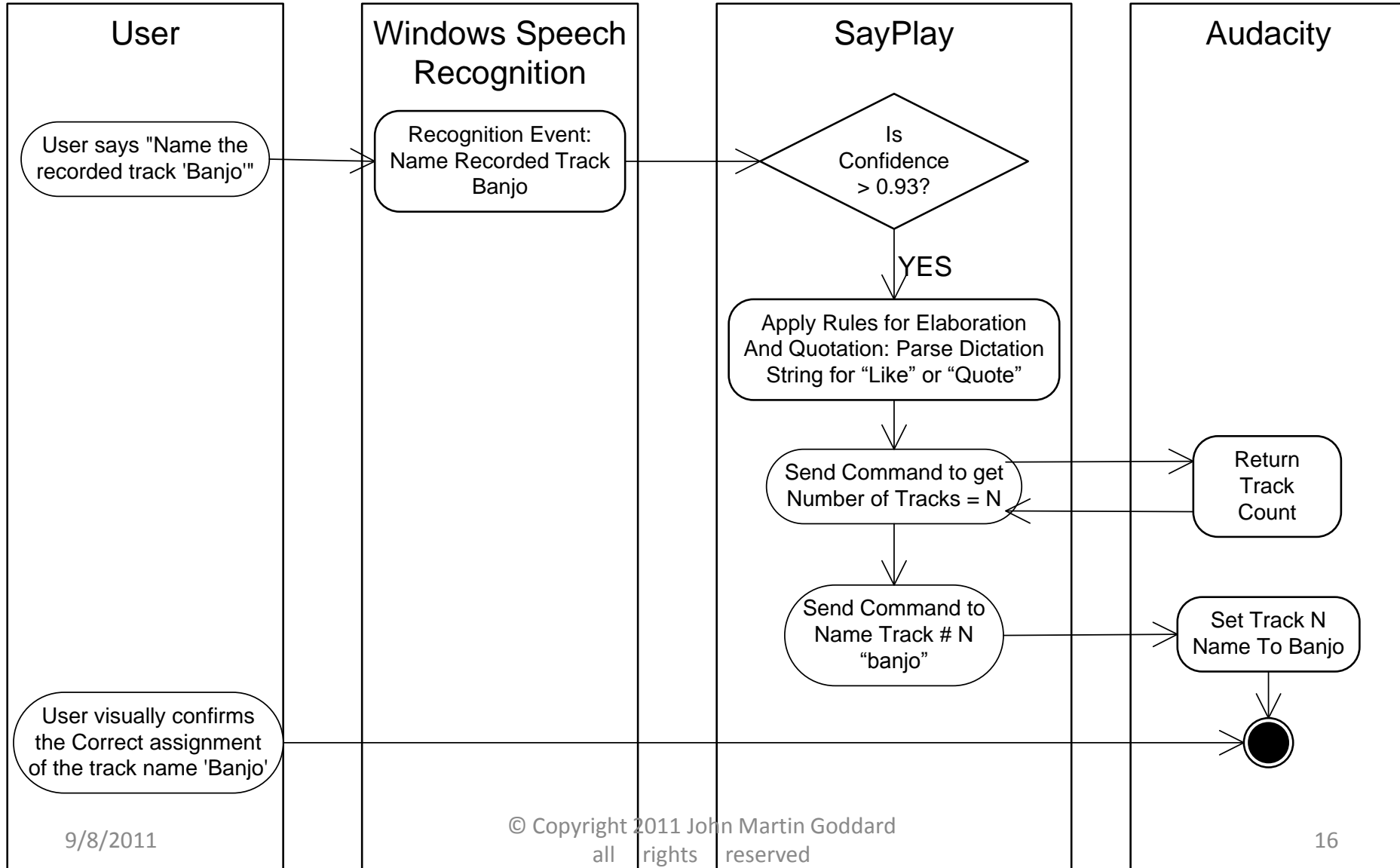
Grammar Structure for Naming Tracks

Creating the Name



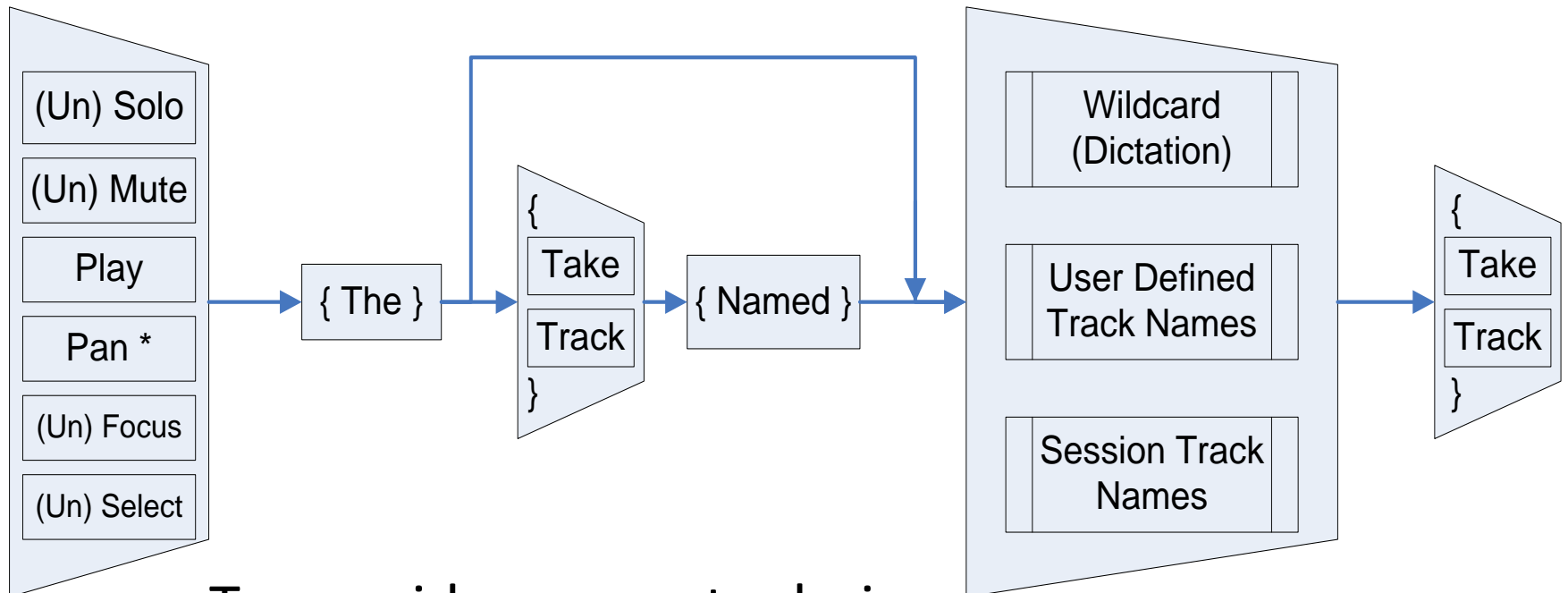
- Trapezoid represents choices
- { } Indicates optional word(s)

Steps to name a track "Banjo"



Grammar Structure for Named Tracks

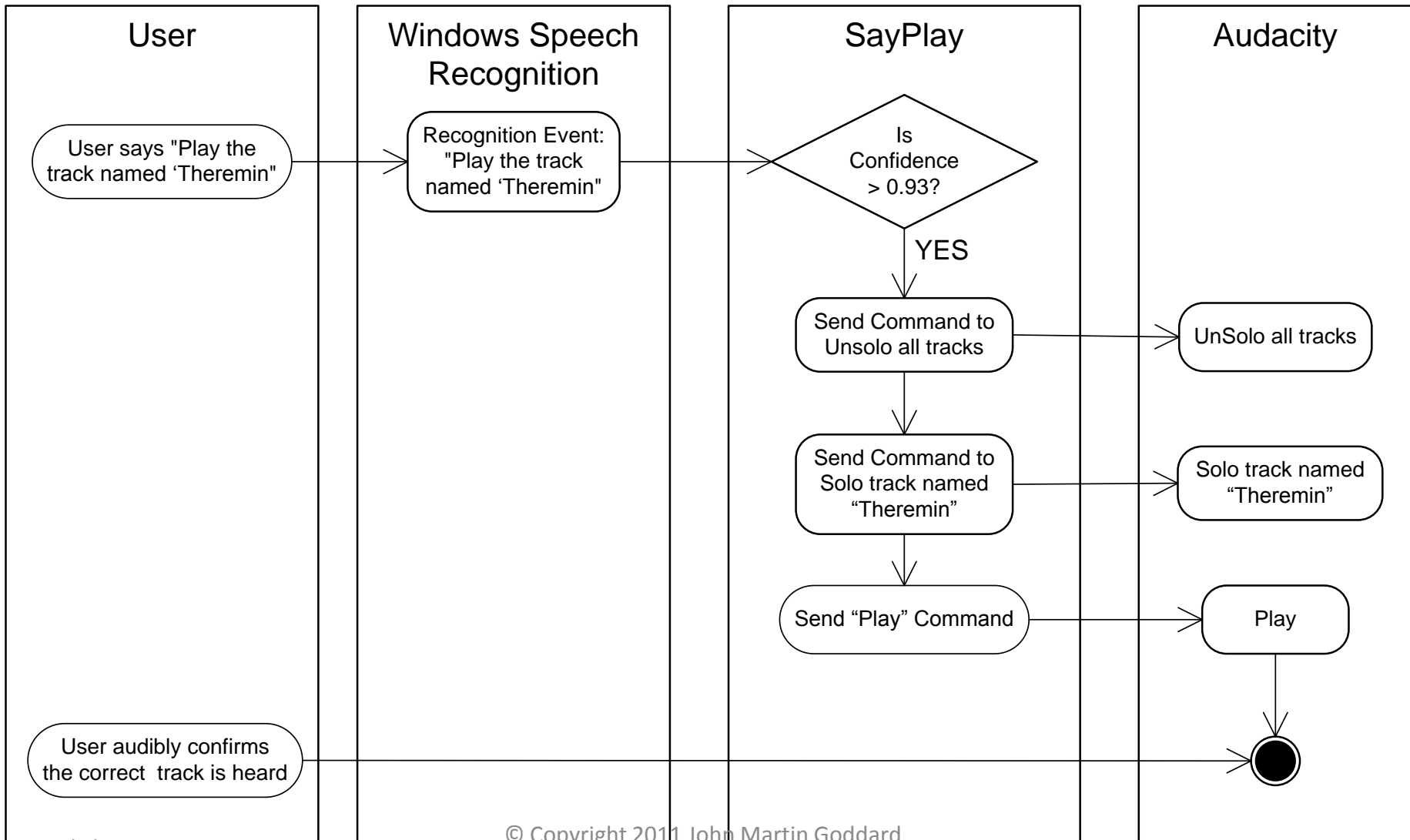
Using the Name



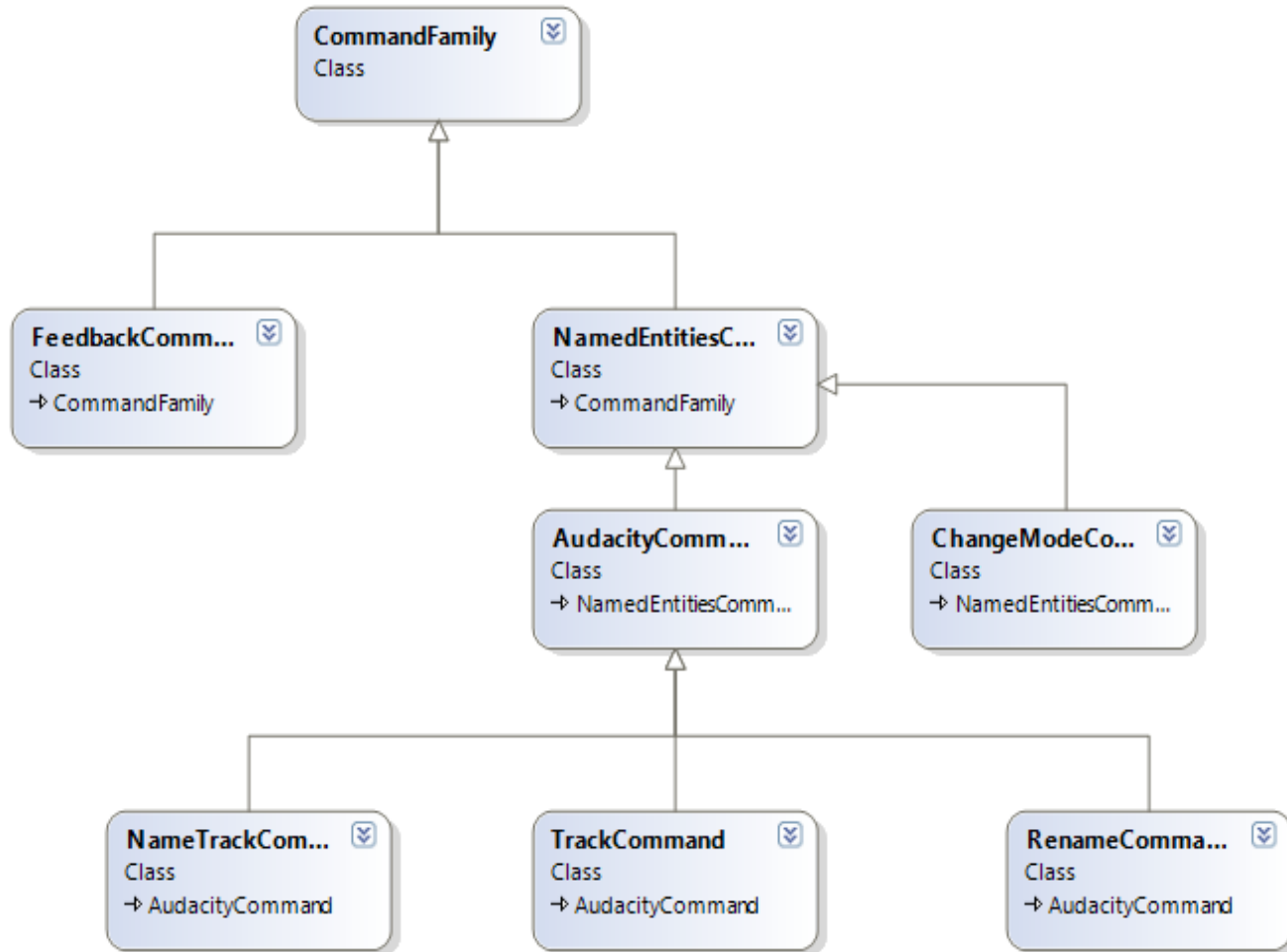
- Trapezoid represents choices
- { } Indicates optional word(s)

* Pan command appended with “Medium Left”, “Hard Right” Etc.

Steps to Play a Named Track



“CommandFamily” hierarchy



Summary of Contributions

- Voice Commands to support recording workflow
- Apparatus to allow easy experiments and interpretation of events.
- Flexible Commands to Name & Refer by name
- Techniques to improve accuracy of tricky names
 - Elaboration and Quotation
 - Spelling it out
 - Adding word to dictionary
 - Preventing Dictation of Confused Words
 - Add names to grammar (so dictation speech recognition isn't required)

Future Work:

- Load all tracks names into the grammar structure in the Speech Recognition Engine
- Add new names to Speech Dictionary (whenever loading them into the grammar)
- Prevent recognition of confused words (Such as when user says “Wrong” then repeats the naming command. The wrong word recognized should be prevented from subsequent recognition.)

More Future Work on Voice Commands to Control Recording Sessions

- Name song sections for navigating in time
 - “Name this section ‘chorus’, or ‘verse’, or ‘hook’”
 - “Play the Chorus”, or “Jump to the solo”
- Reference multiple objects in one command
 - “Mute the piano, bass and backing vocal tracks”
- Rename Commands with “This Means That”
 - To help a person remember exact command phrases
 - “Computer, please note that the command ‘Audition’ means the same as the play command”

Other Possible Future Work

- Naming tasks, processes, or searches for voice commands to complete user-defined multi-step commands.
- Transpose this work to
 - Personal note taker or PDA
 - Video recording/playback
 - General Voice User Interface effectiveness