

# Voice Commands to Control Recording Sessions

## Handout 3: Techniques for Naming Tracks, & Referring to them by name

*3 Techniques in Assigning Names:  
Spell it out, Quote/unquote, and Elaboration*

*3 Techniques in Using Names:  
Load into Grammar, Add to Dictionary, and prevent  
recognition of confusing words*

John “Marty” Goddard

# Name Recognition Failures

- **Wrong Name**: track name is misrecognized
- **“Breath After”**: WSR misrecognizes a dysfluency following a recognized word, whenever an optional additional word is possible. Ex: “Wrong *and*”
- Name recognition failures were the most prevalent of all failure modes during testing.

# Overview

- Improving Accuracy of Name Commands
  - 3 Techniques in Assigning Names
    - Spell it out, Quote/unquote, and Elaboration
  - 3 Techniques in Using Names
    - Load into Grammar, Add to Dictionary, and prevent recognition of confusing words
- Rules applied to Dictation Recognition text string
- Future work and ideas for improvement

**Name Commands:** It was found that “Wrong Name” commands were most prevalent

- Naming tracks (future: takes, sections, markers)  
Ex: “Name this track ‘kazoo’ “
- Adjusting track parameters: Refer to tracks by assigned name, to control settings (mute, solo , pan, gain, etc.)

Ex: “Play the kazoo track”, “Mute the ‘violin’ track”

# Problem With Dictation Speech Recognition: Names from “out of the blue”

- Random names (even made up ones), require large vocabulary dictation speech recognition
- Techniques were developed to improve accuracy assigning names and using names
- Once a name is assigned, it can still be misrecognized whenever using it.

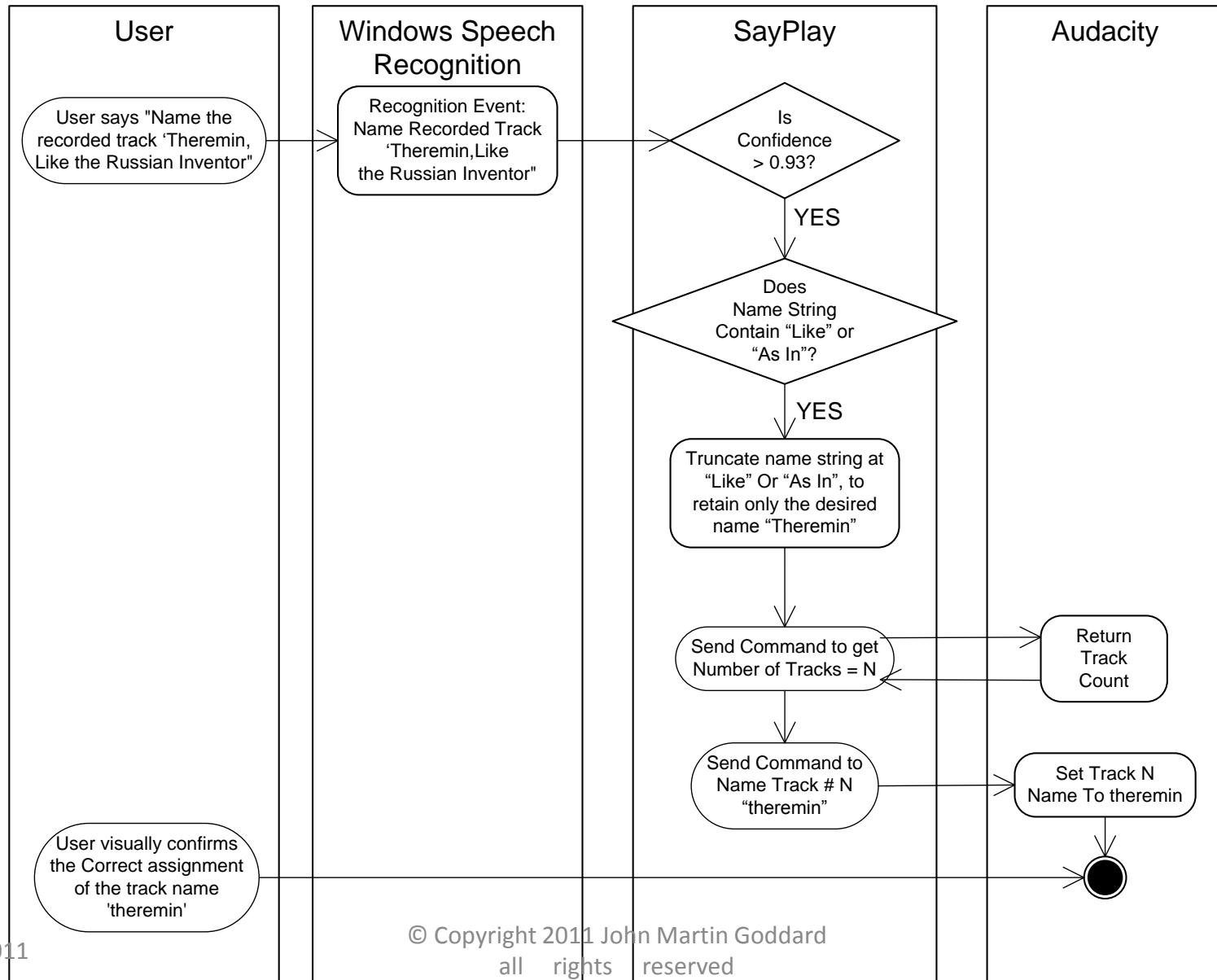
# 3 Techniques Developed to Improve Assigning Names

- **1. Elaboration:** Use “Like” or “As In” to add words to a phrase about, or containing the desired name. Ex: “Name this track ‘bass’ as in bass guitar”.
- **2. Quotation:** Enable a quoted phrase to be the name. Ex: “Name this track quote ‘scream like a banshee’ unquote”.
- **3. Spell it:** “Name this track *spelled* W, O, W”.

# 1. Elaboration with “Like” or “As In”

- Added details about a name provide the speech recognition engine with extra words to analyze
- N-Gram analysis provides better recognition accuracy when extra words add distinction
- Examples: “Name this track...
- ...‘bass’ like bass guitar”
- ...‘Theremin’ like science fiction movie sounds”
- ...‘Alto Sax’ like the saxophone”

# Activity Diagram of Elaboration





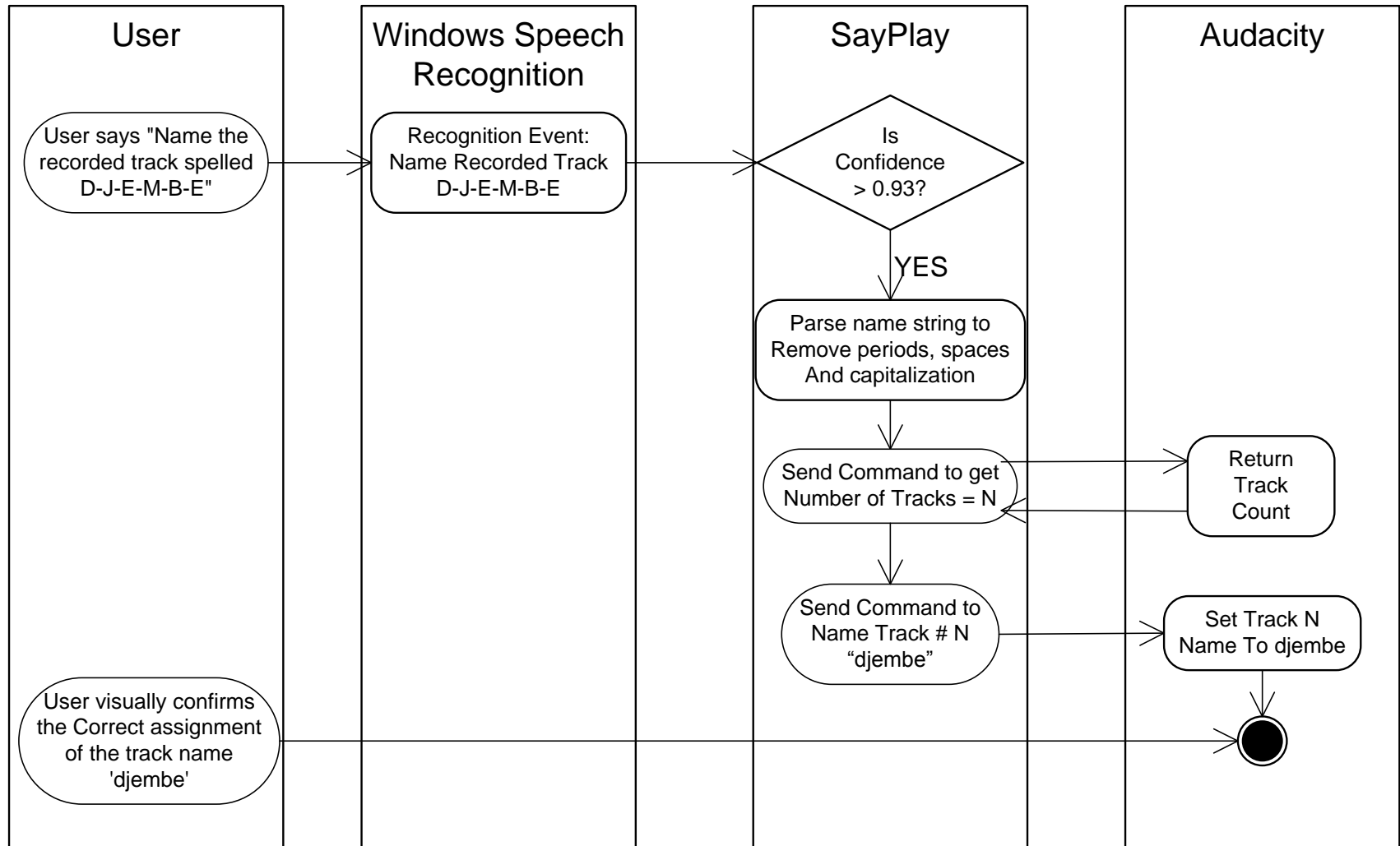
## 2. Quote/Unquote & “As Follows:”

- User can override the truncation of “Like” and capture an entire phrase to use as the name  
Ex: “Name this track ‘quote’ scream *like* a banshee ‘unquote’.”
- Or, user can say: “Name this track as follows...” and then give the name.  
Ex: “Name this track as follows: ‘scream *like* a banshee’.”

# 3. Spell the name

- User can spell out name when all else fails  
Ex: “Name this track *spelled* W, O, W. ”
- It works, but the user will likely lose patience having to spell it out whenever referring to names already assigned (e.g. it should *learn*).
- Instead, the correct name should be added to the grammar and/or the speech dictionary...

# Activity Diagram of Spelling out a track name



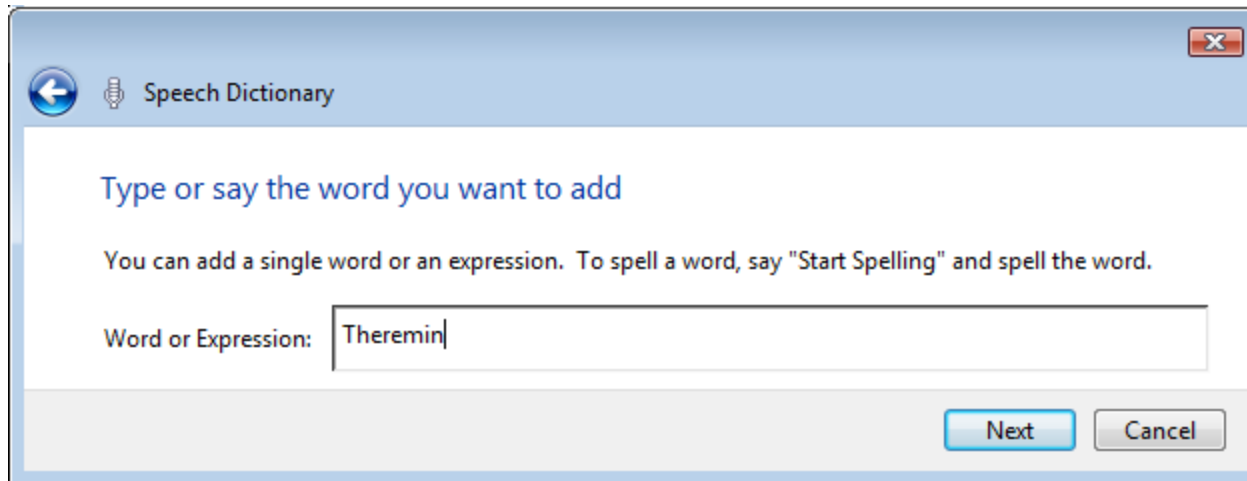
# 3 Techniques to Refer to a Named Item, Once the Name is Assigned

(In addition to using previous 3 techniques)

1. Add the name to Windows Speech Dictionary
2. Prevent recognition of confused names
3. Add current track names into the grammar loaded into the Speech Recognition Engine

# 1. Add Name to Dictionary

- “Open Speech Dictionary”
- “Add a new word”
  - Essential for words not in dictionary
  - User can record a spoken example





## Add your word

To add your word to the Speech Dictionary, click Finish. To ensure your word is dictated properly, you may choose from the following options:

Therein

### Pronunciation

If the pronunciation of your word is not obvious, you can record a pronunciation to improve Speech Recognition accuracy.

**Record a pronunciation upon Finish**

### Capitalization

- This word is always capitalized (e.g., Detroit)**
- This word is sometimes capitalized (e.g., Summer or summer)**

**I want to make more modifications to the Speech Dictionary upon Finish**

Finish

Cancel

## 2. Prevent Recognition of Confused Names

- “Prevent a word from being dictated”
  - Results are mixed (for example: “Theremin” )
    - Preventing “Sarandon” stops it from being recognized
    - But, new incorrect result is “Salmon”, “Thurmon” or “Fairman”
  - Should be applied to the recent incorrect result
    - “Name this track ‘Theremin’ results in ‘Sarandon’” then after user says ‘wrong name’, it should automatically prevent ‘Sarandon’ from subsequently being recognized.
  - Method to automatically do this (in software) is undocumented & unknown (suggestions are welcome)



Listening

## Prevent a word from being dictated

You should only exclude words that you never dictate or that are the cause of frequent dictation mistakes. To spell a word, say "Start Spelling" and spell the word.

Word:



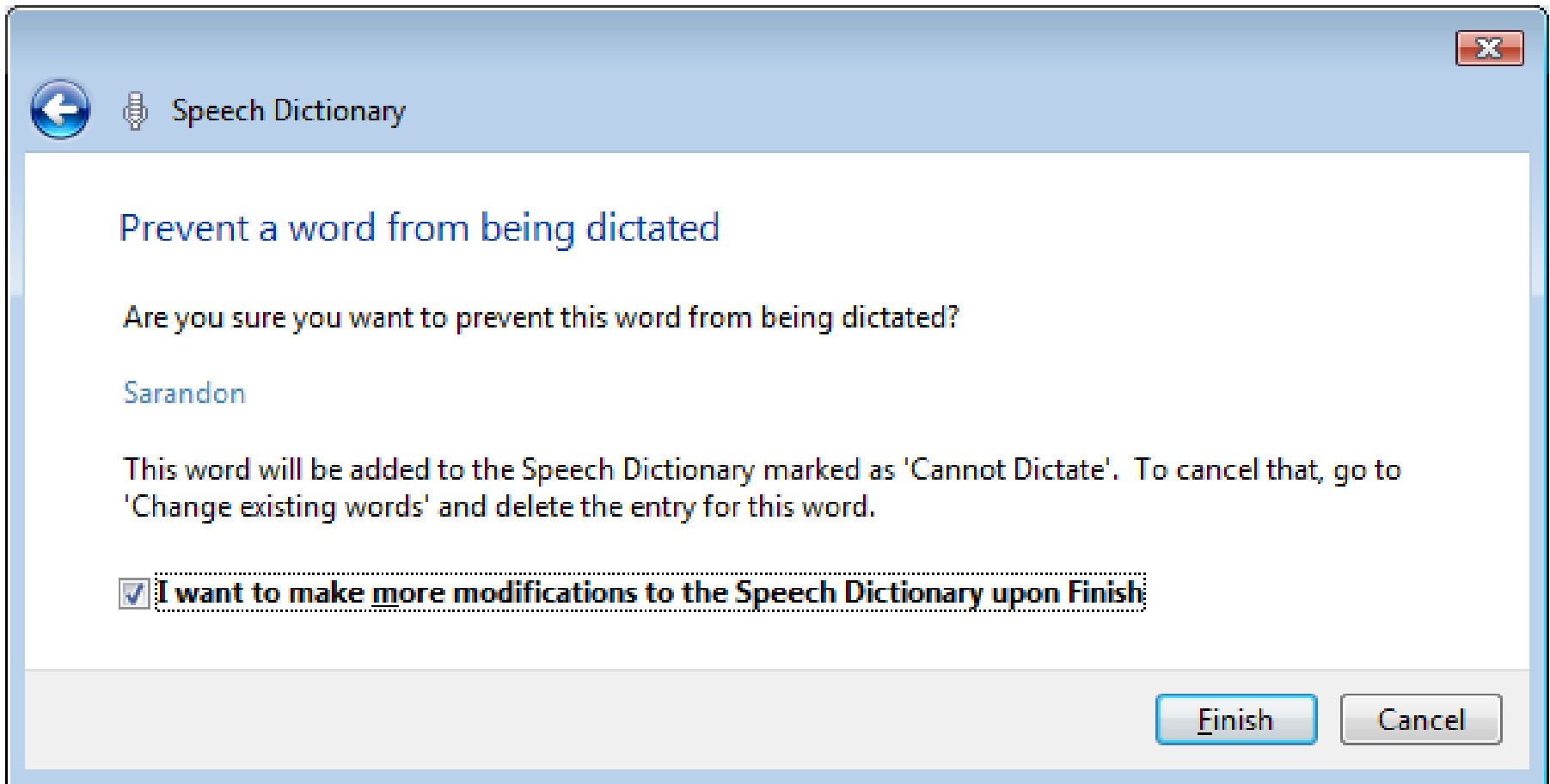
Note: You may not enter spaces, numbers, or symbols here.  
(For example, you may not enter "Los Angeles".)

Make sure that the word you enter has the same capitalization as the word you want to exclude from dictation. If the word is sometimes capitalized, enter it in all lower case.

Next

Cancel





I wish this could be done automatically (programmatically) whenever the user says “Wrong” and must repeat a name assignment, or command.

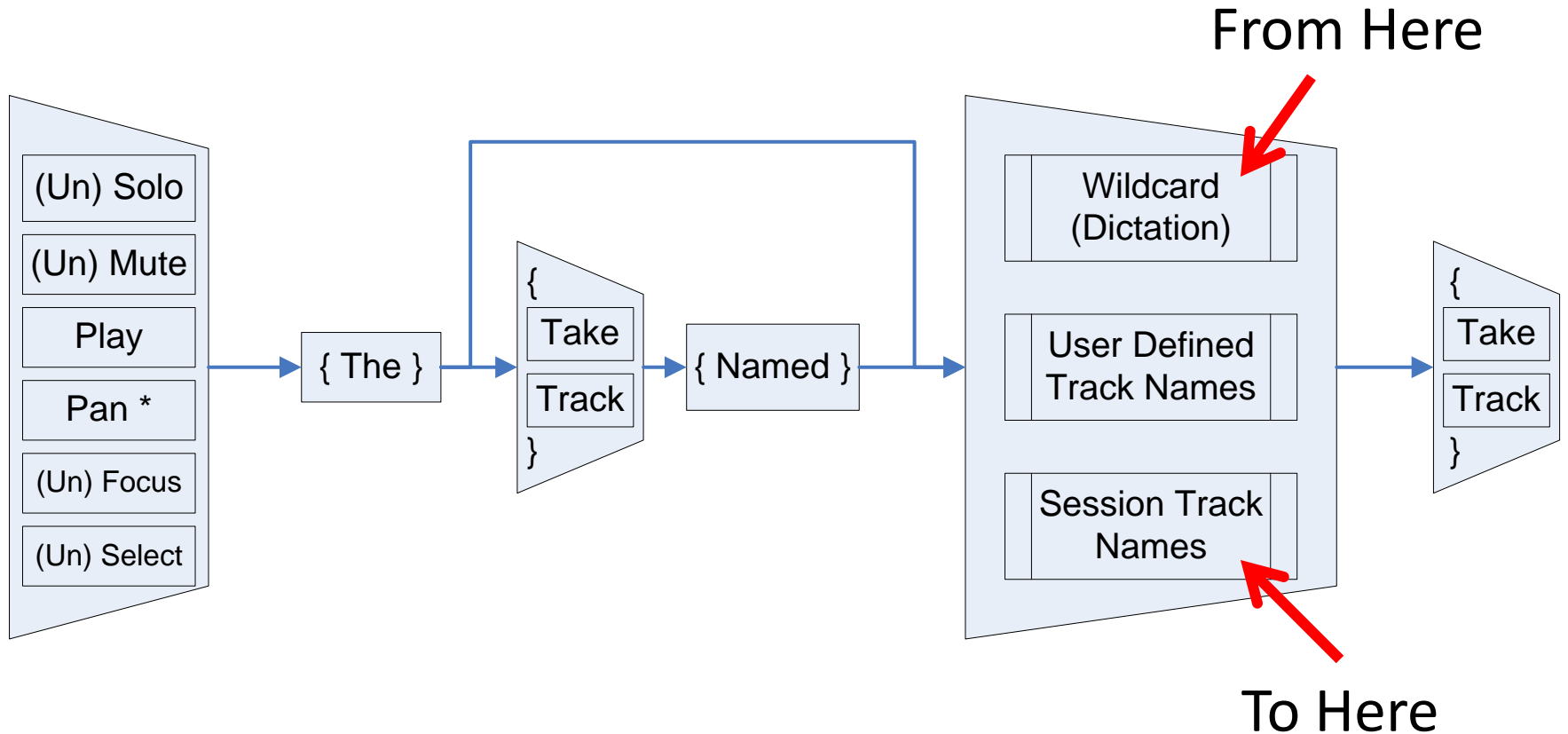
# 3. Loading Names into Grammar

“Computer, please refresh the session”

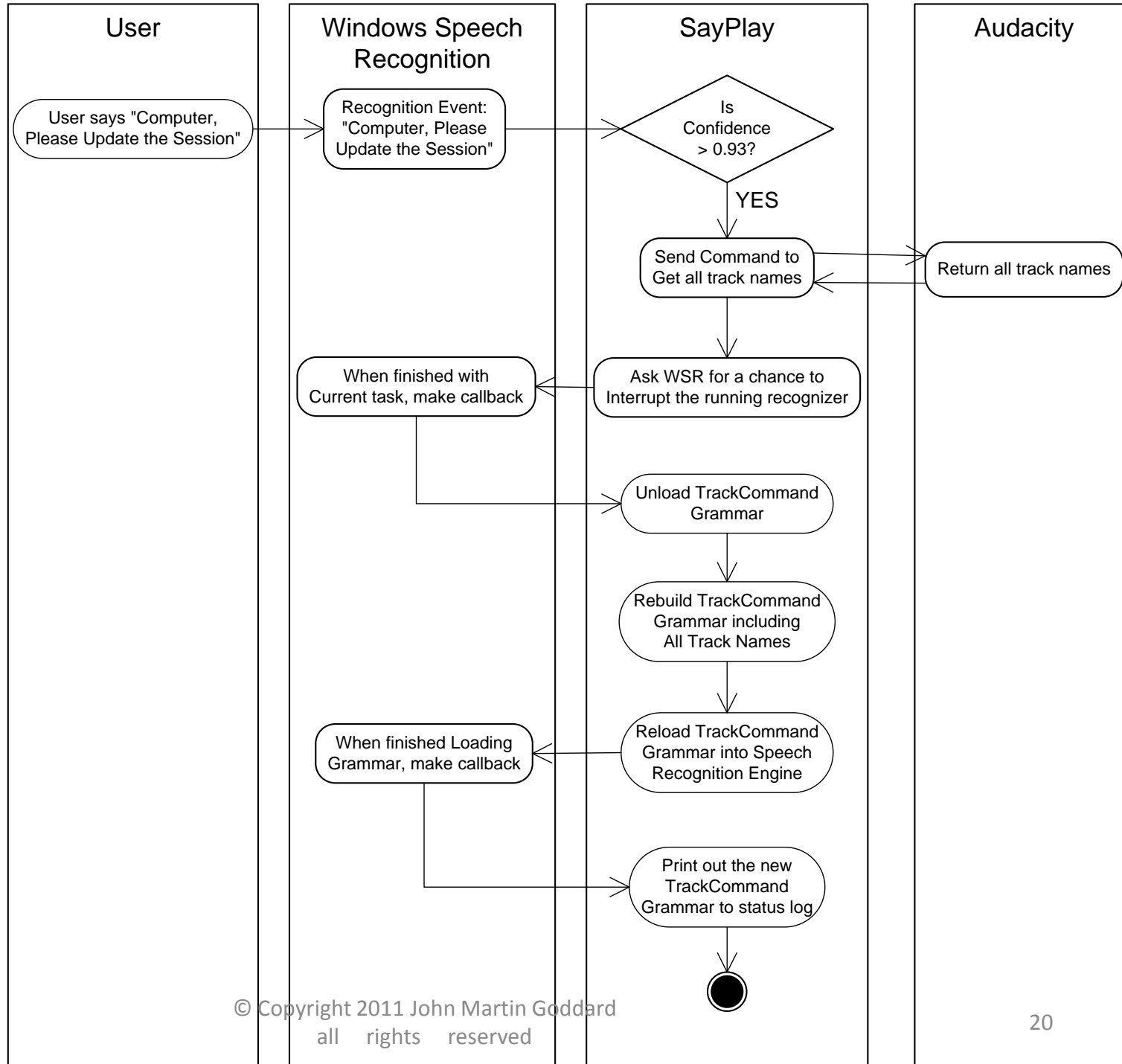
- Get the list of all track names from Audacity
- Register a request with the Speech Recognition Engine, for a time slot, which generates a callback
- Upon the callback:
  - Unload the Track Commands Grammar
  - Construct a new grammar including the track names
  - Asynchronously reload the Track Commands Grammar
- Upon callback for Asynchronous Load Complete
  - Print out the new Track Command grammar
- Now, track names can be changed again, but another refresh will be necessary

# Loading Names into Grammar

## Moves Recognition Duty...



# Steps to Load Names into Grammar



# Rules applied to Text Strings returned as Track Names from Dictation SR

- ❖ Truncate “Like” or “As In” and everything after, and keep rest as name (if present in string).
- ❖ Delete “Quote” and “Unquote”, or “As Follows”, from name string, and retain remainder as name (even if it contains “Like” or “As In”)
- ❖ Delete “I”, “Ah”, “And”, or “A” whenever appearing last in a name string (solves “Breath After” bug)
- ❖ Delete a leading “The” because this is part of the command, and will prevent name recognition.

# Creating and Using Names

- Explorations with invented names
  - Sticklavier
  - Gambales
- Surprisingly Recognized Names
  - Whoosh, Vox, Improv
- Proper Names are favored over other words:
  - Shanti, Wei, Kazue, Orion, Sarandon, Peres, etc.
- Names Commonly Mistaken
  - Violins mistaken as violence
  - Bass mistaken as Base
  - Alto Sax mistaken as Alta Sachs

# Further Examination of Assigning Names, and using them

- Names tested: Wow, While, Gambales, Crotales, Theremin, Kazoo.
- Other names to experiment with:
  - names of persons
  - Name with a numbered offset (Bass Two, Bass 3)
  - Foreign language words and names
  - Names with a time of day

# Future Work & Ideas for Improvement:

Always load track names into Grammar whenever opening an existing session in Audacity

Automatically (Programmatically ):

- 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.)

(Because these currently must be user-initiated)



# 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”
- Reference multiple objects in one command
  - “Mute the piano, bass and backing vocal tracks”
- Allow commands to be renamed using “This Means That”
  - “Computer, please rename the ‘Play’ command ‘Audition’