Voice Commands to Control Recording Sessions

Handout 2: Synopsis of Experiments

Learning/Improvement
Basic Command Phrases
Tricky Names
Homophone Experiments:
Elaboration and Loading Names into Grammar

Thesis in partial fulfillment of the degree
Master of Science in Computer Engineering
John "Marty" Goddard

Why we need experiments

- Complexity of workflow

  Record then playback of specific tracks

  Save or delete

  Assign "best" or "alternate" status

- Somewhat unpredictable WSR results

  Performance generally increases

  Name Confusion

  Wrong context (user error)

Exploration Strategy

(due to complicated context, and imperfect recognition)

- Measure accuracy

- Issue a command (ideally >25 times)

- Calculate success rate

- Develop an improvement

- Test for actual performance increase

- Develop another improvement

- Continue to test for actual performance increase

Basic Measurement Method

Step 1: Issue Voice Command

Step 2: Indicate whether or not command executes
- Pass/Fail determination based on inspection
- On success, verbally state: "Pass" or "Correct"
- On failure, say: "Wrong" or "Fail"

Command Events & Results are written to log file

- Comments can be added after each event
- Or, a comment can be announced any time: "Comment…"

Log File excerpt: Theremin and Tapping

Command recognized by WSR engine

Confidence value returned by WSR engine

Pass/Fail Determination

```
Pan the theremin soft left        0.9680607       Correct
Pan the theremin soft right       0.9697677       Correct
Pan the theremin center 0.999    Comment this version of the code
    doesn't support pending into the center
Mute the track named theremin     0.9338306       Correct
Solo the track named rhythm guitar right          0.9492097
        Correct
Mute the taping track   0.9552542       Wrong it keeps getting it
    confused with taping
Solo the track named tapping      0.9575906       Correct
UnSolo the taping track 0.9783866           Comment now that time I said
    the wrong name
Computer please Refresh the session       0.9883651
UnMute the theremin track         0.9835022
UnSolo the tapping      0.9790956
UnSolo rhythm guitar right        0.9636887
Mute the theremin       0.9785874
```

User Comments

Summary of Activities: We have already successfully named a track "Theremin" and we are referring to it (Pan, Solo, Mute, etc.) and to other tracks. N.B. The "Tapping" track is misrecognized as "Taping".

Compiling the experimental results:

SayPlay Software Version:	11/29/2010 18:24		140	49		79	3	0	0	0	0	9
SayPlayLogFile_1910892419			Correct	cumulative correct	wrong:	wrong name	low confidence	compound comnd failure	wrong phrasing	timeout violation	false positive	true rejection
Focus the clapp track	0.9625631	low confidence	0	0		0	1	0	0	0	0	0
Focus the clapp track	0.9309105	Correct	1	1		0	0	0	0	0	0	0
Name this track fairman	0.9274955	wrong name	0	1		1	0	0	0	0	0	0
Name this track fairman	0.9552792	wrong name	0	1		1	0	0	0	0	0	0
Name this track famine like the instruments	0.8906806	wrong name	0	1		1	0	0	0	0	0	0
Name this track fairman like the instrument	0.9525561	wrong name	0	1		1	0	0	0	0	0	0
Name this track if they remain like electronic instruments	0.9558795	wrong name	0	1		1	0	0	0	0	0	0
Name this track fairman like the science fiction movie sound	0.9741226	wrong name	0	1		1	0	0	0	0	0	0
Name this track fairman like the Russian inventor	0.9547242	wrong name	0	1		1	0	0	0	0	0	0
Name this track theremin like the science fiction movie sound	0.9581242	Correct	1	2		0	0	0	0	0	0	0
Name this track the track fairman like the	0.6480187	true rejection	0	2		0	0	0	0	0	0	1
Name this track thurman like the musical instrument	0.9326818	wrong name	0	2		1	0	0	0	0	0	0
Name this track fairman like the musical instruments	0.9514031	wrong name	0	2		1	0	0	0	0	0	0
Name this track fairman like the musical instrument	0.9565645	wrong name	0	2		1	0	0	0	0	0	0
Name this track thurmond like the Russian inventor	0.8099763	wrong name	0	2		1	0	0	0	0	0	0
Name this track theremin like the Russian inventor	0.9560577	Correct	1	3		0	0	0	0	0	0	0

LogFile of Experiment

Software Version

Correct Tally

Errors by Error Type

SayPlay Software Version:	11/29/2010 18:24	140	49				79	3	0	0	0	0	9
SayPlayLogFile_1910892419			Correct	cumulative correct		wrong:	wrong name	low confidence	compound comnd failure	wrong phrasing	timeout violation	false positive	true rejection
Focus the clapp track	0.9625631	low confidence	0	0			0	1	0	0	0	0	0
Focus the clapp track	0.9309105	Correct	1	1			0	0	0	0	0	0	0
Name this track fairman	0.9274955	wrong name	0	1			1	0	0	0	0	0	0
Name this track fairman	0.9552792	wrong name	0	1			1	0	0	0	0	0	0
Name this track famine like the instruments	0.8906806	wrong name	0	1			1	0	0	0	0	0	0
Name this track fairman like the instrument	0.9525561	wrong name	0	1			1	0	0	0	0	0	0
Name this track if they remain like electronic instruments	0.9558795	wrong name	0	1			1	0	0	0	0	0	0
Name this track fairman like the science fiction movie sound	0.9741226	wrong name	0	1			1	0	0	0	0	0	0
Name this track fairman like the Russian inventor	0.9547242	wrong name	0	1			1	0	0	0	0	0	0
Name this track theremin like the science fiction movie sound	0.95812	rrect	1										
Name this track the track fairman like the													
Name this track fairman like													

Failure Classification

Attempts to Name a track "Theremin" with various uses of Elaboration

Types of Recognition Failures
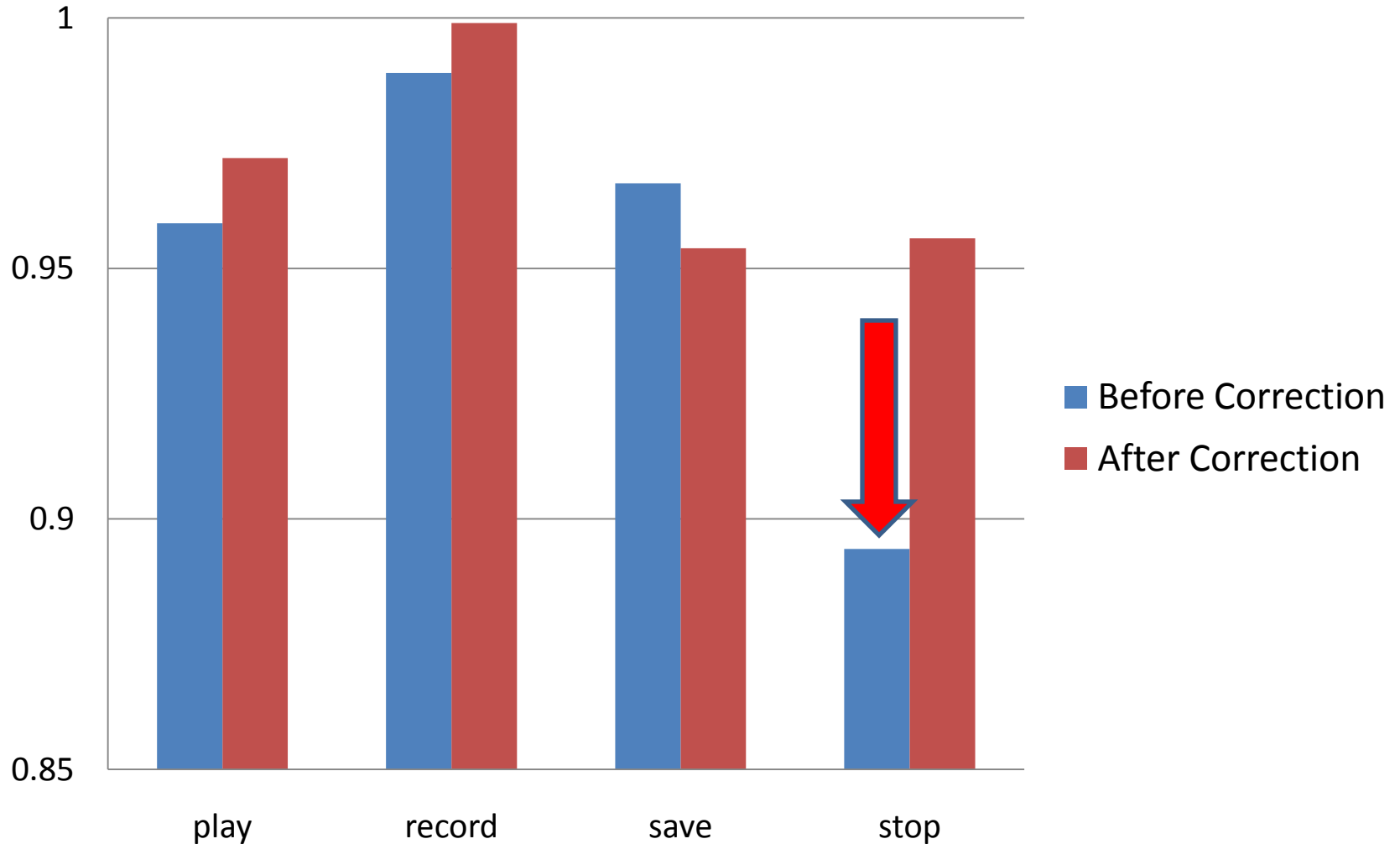
- **<u>Wrong Name</u>:** track name is misrecognized
- **<u>Wrong phrasing</u>:** User error, incorrect wording
- **<u>Low Confidence</u>:** Correctly recognized but value of confidence (returned by WSR) is below threshold
- **<u>Timeout</u>:** Long pause in speaking truncates phrase
- **<u>False positive</u>:** Utterance misinterpreted as a command, with confidence above threshold
- **<u>True rejection</u>:** A non-command is misrecognized, but confidence is below threshold, so it is justly rejected
- **<u>"Breath After"</u>:** WSR misrecognizes a dysfluency following a recognized word. Possible when an *optional additional* word is allowed. Ex: "Wrong <u>*and*</u>"

Results on Basic Commands

- Variation of Confidence Values:

  Average Confidence per Command

  - Play: 0.951243
  - Stop: 0.894
  - Pause: 0.936
  - Record: 0.989
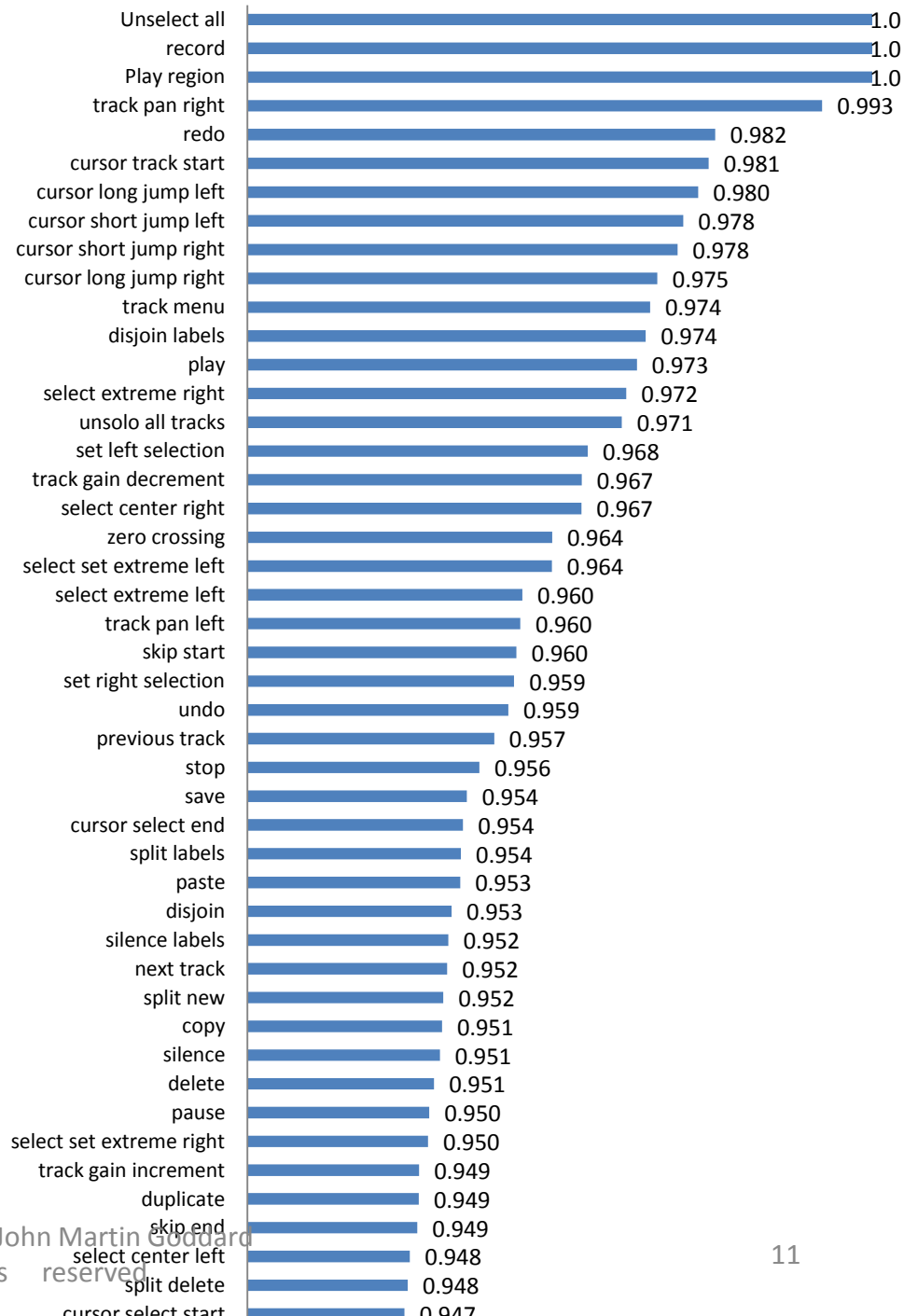 - Save: 0.967
  - Set (Left/Right) Selection: 0.916

Confidence before and after mic setup

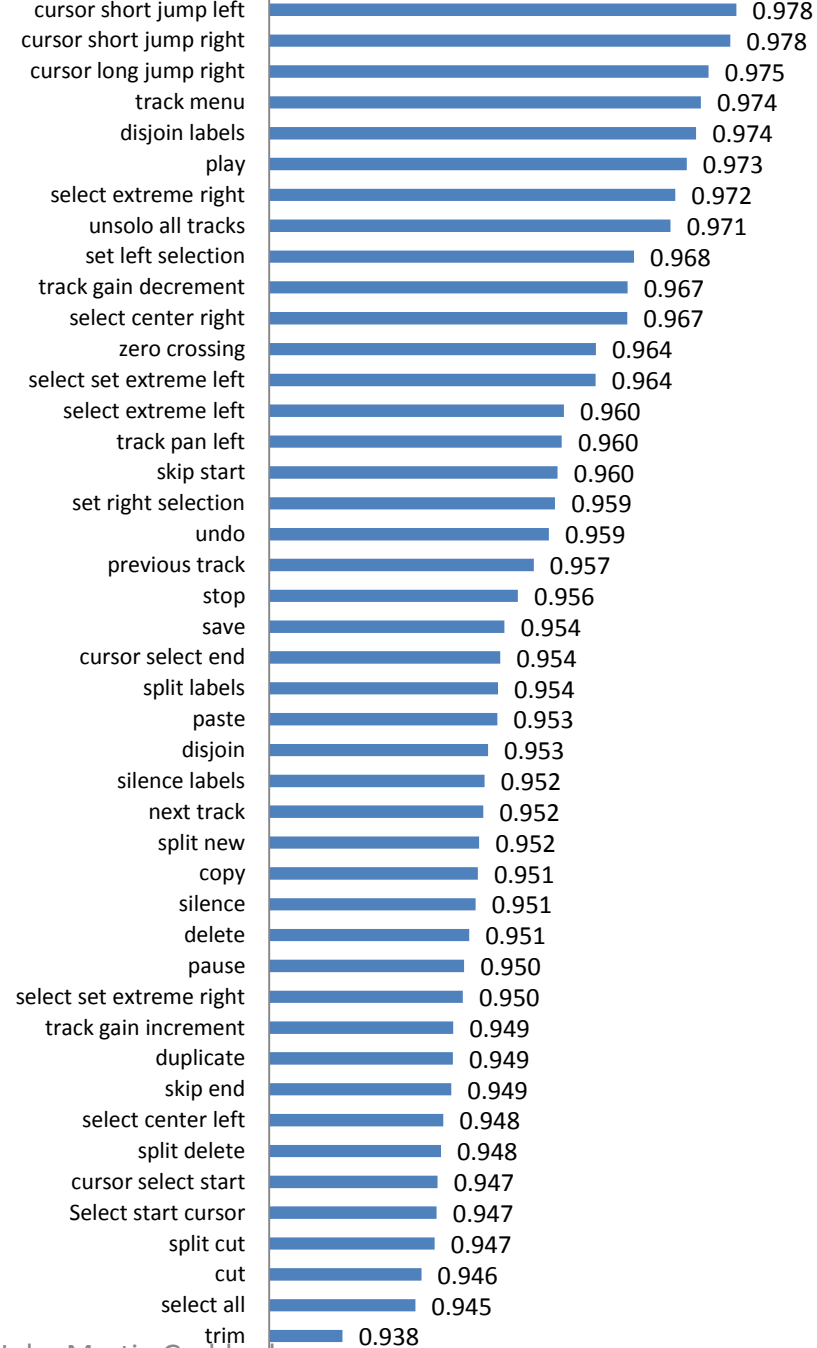## Average "Confidence" returned by WSR for basic commands

Single Phrase Commands:

Confidence value returned by Speech Recognition Engine

Command	Value
Unselect all	1.0
record	1.0
Play region	1.0
track pan right	0.993
redo	0.982
cursor track start	0.981
cursor long jump left	0.980
cursor short jump left	0.978
cursor short jump right	0.978
cursor long jump right	0.975
track menu	0.974
disjoin labels	0.974
play	0.973
select extreme right	0.972
unsolo all tracks	0.971
set left selection	0.968
track gain decrement	0.967
select center right	0.967
zero crossing	0.964
select set extreme left	0.964
select extreme left	0.960
track pan left	0.960
skip start	0.960
set right selection	0.959
undo	0.959
previous track	0.957
stop	0.956
save	0.954
cursor select end	0.954
split labels	0.954
paste	0.953
disjoin	0.953
silence labels	0.952
next track	0.952
split new	0.952
copy	0.951
silence	0.951
delete	0.951
pause	0.950
select set extreme right	0.950
track gain increment	0.949
duplicate	0.949
skip end	0.949
select center left	0.948
split delete	0.948
cursor select start	0.947

Single Phrase Commands:

Confidence value returned by Speech Recognition Engine (cont.)

Command	Value
cursor short jump left	0.978
cursor short jump right	0.978
cursor long jump right	0.975
track menu	0.974
disjoin labels	0.974
play	0.973
select extreme right	0.972
unsolo all tracks	0.971
set left selection	0.968
track gain decrement	0.967
select center right	0.967
zero crossing	0.964
select set extreme left	0.964
select extreme left	0.960
track pan left	0.960
skip start	0.960
set right selection	0.959
undo	0.959
previous track	0.957
stop	0.956
save	0.954
cursor select end	0.954
split labels	0.954
paste	0.953
disjoin	0.953
silence labels	0.952
next track	0.952
split new	0.952
copy	0.951
silence	0.951
delete	0.951
pause	0.950
select set extreme right	0.950
track gain increment	0.949
duplicate	0.949
skip end	0.949
select center left	0.948
split delete	0.948
cursor select start	0.947
Select start cursor	0.947
split cut	0.947
cut	0.946
select all	0.945
trim	0.938

Experiments with test subjects

- User is asked to perform the following actions in sequence.
    1. Record a Track
    2. Name the track
    3. Record and/or Name another track
    4. Listen to only the first track by Number or Name
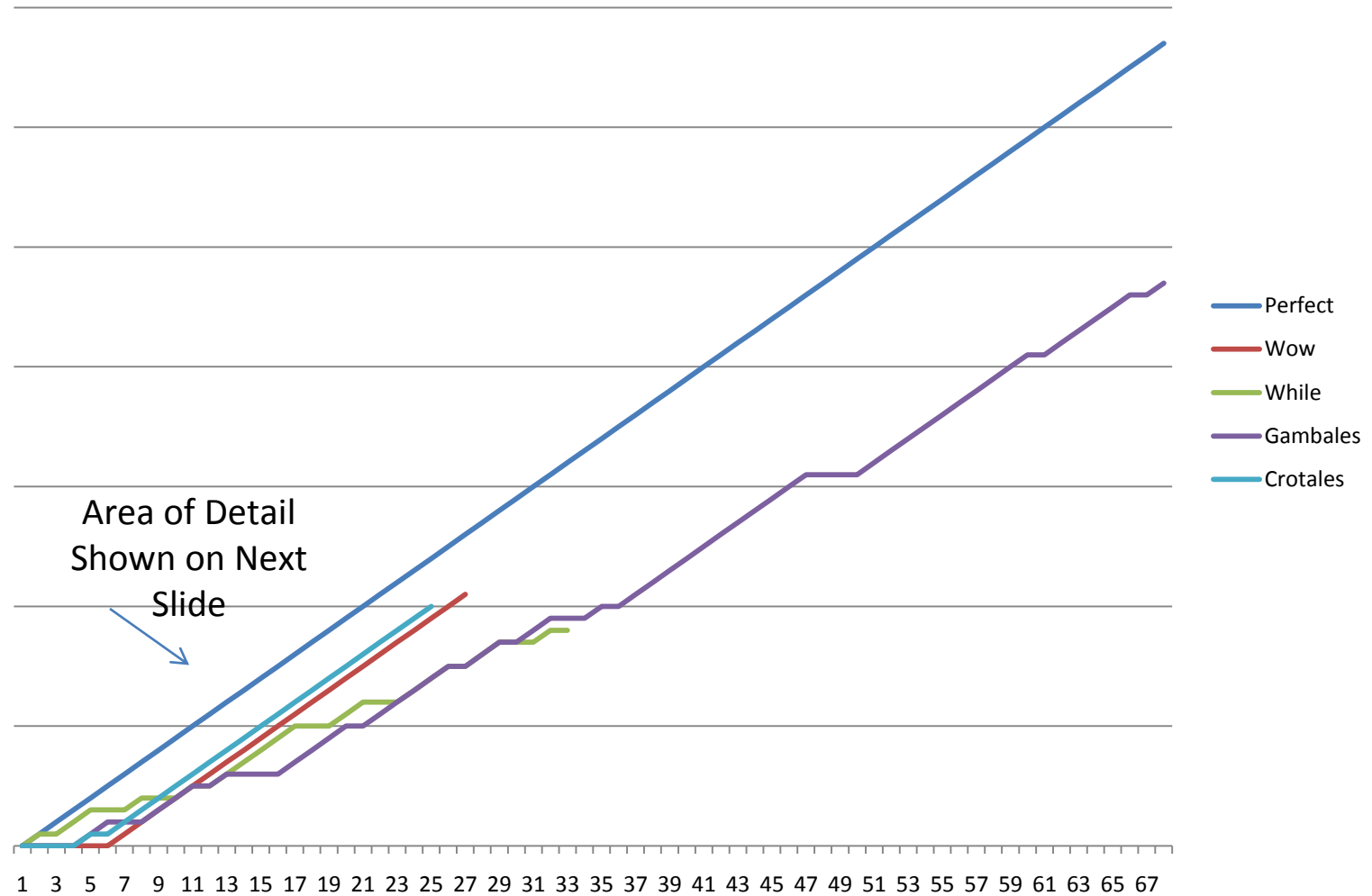   5. Listen to the second track by Number or Name

Learning Performance

63 = "Vocal Scat"

63

55 = "Vocal Scat"

47 = "Violins"

21 = "Vocal Scat"

26 = "Tambourine", "Alto Sax"

Alternating success/failure = "Pass"
misrecognized as "Pause"

7 = "Banjo", "Guitar", "Saxophone"

Legend:
- Perfect
- A
- B
- C
- D
- E
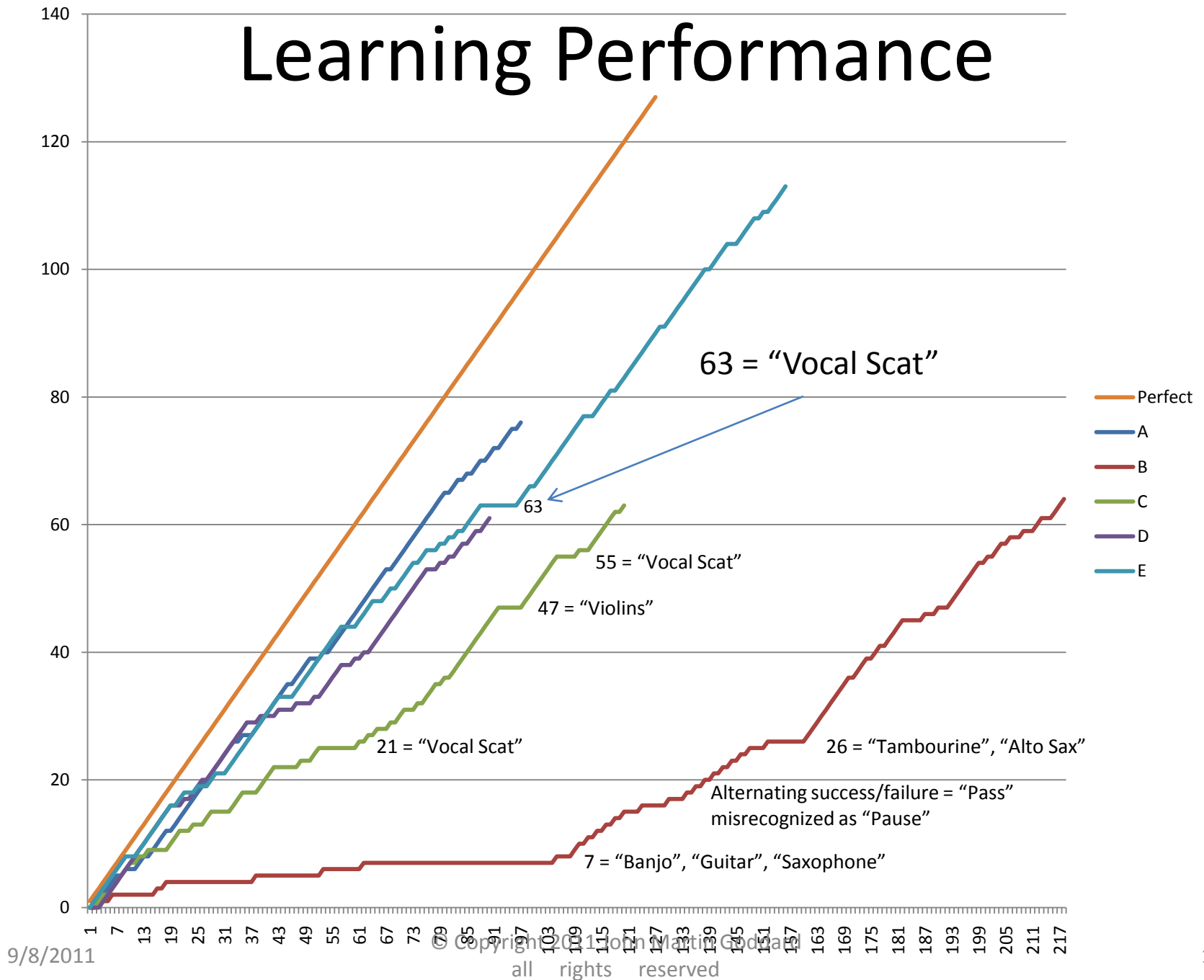
Creating and Using Names

- Names Commonly Mistaken
  - Violins mistaken for violence
  - Bass mistaken for Base
  - Alto Sax mistaken for Alta Sachs
  - Wow mistaken for While

- Explorations with invented names
  - Sticklavier
  - Gambales

- Surprisingly Recognized Names
  - Whoosh, Vox, Improv

9/8/2011

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

9/8/2011

16

Speech Dictionary Results
(single test subject)

Area of Detail
Shown on Next
Slide

Legend:
- Perfect
- Wow
- While
- Gambales
- Crotales

X-axis: 1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67

Speech Dictionary (Detail)

Legend:
- Perfect
- Wow
- While
- Gambales
- Crotales

Zero Values at start are <u>before</u> adding word to Speech Dictionary

x-axis: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
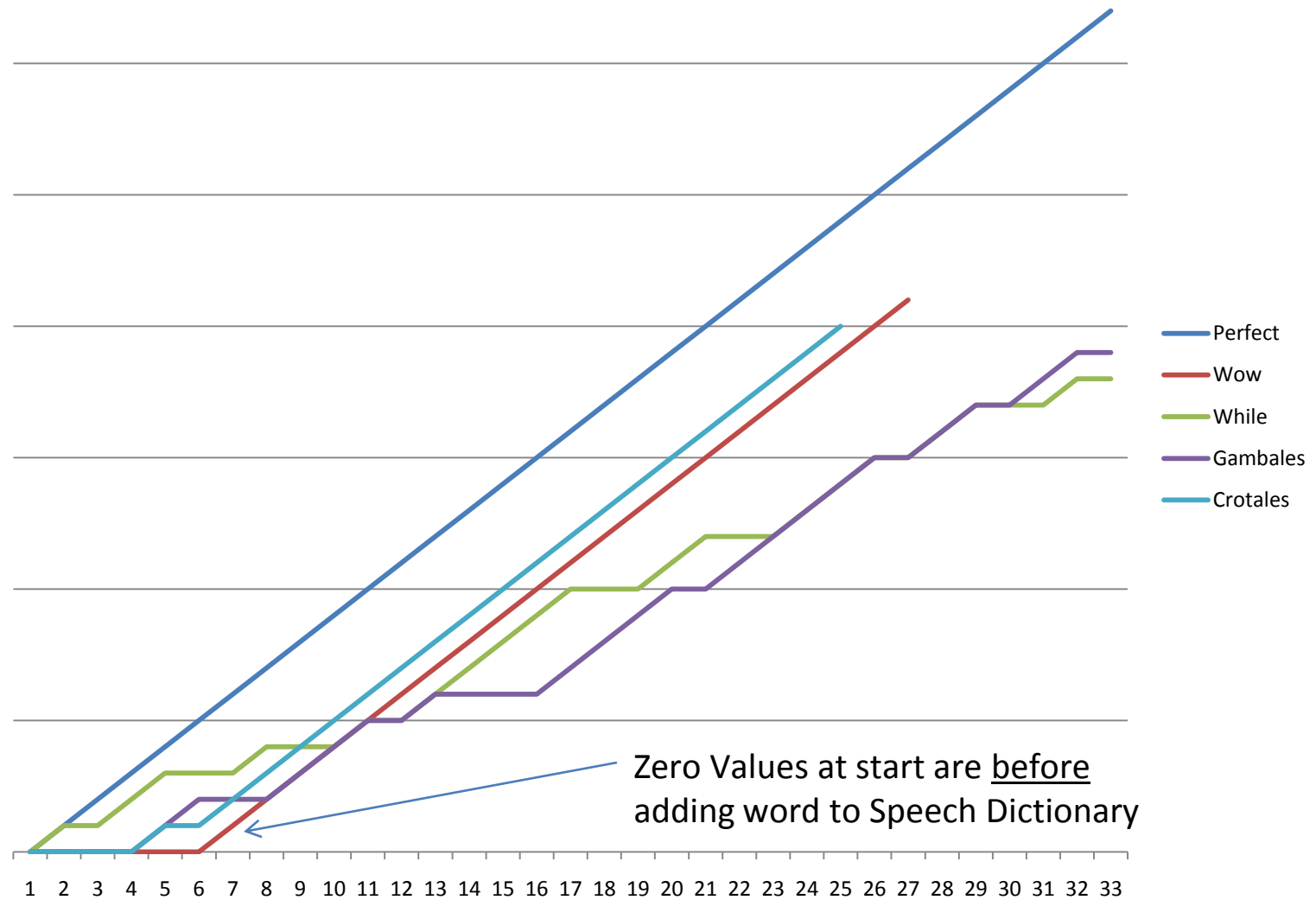
Before and After…

- … adding a word to the Speech Dictionary
  - Theremin (confused with fairman and salmon)
  - Gambales (an invented word)

- … preventing a word from being dictated
  - prevented "fairman", confusion became salmon
  - prevented "salmon", performance increased

- … adding track names to the Loaded Grammar
  - Wow and While were still confused

Before/After Experiments
To show whether techniques are effective

- Method:
  - Setup and train a fresh install of the Microsoft OS
  - Make SayPlay and Audacity work with standard names (to verify basic functionality)
  - Run trials using "tricky" track names
- Repeat experiment, this time using technique:
  - Adding the new name to the dictionary
  - Masking other names it's commonly confused with
  - Include the name in the WSR Grammar (instead of relying on dictation speech recognition for the name).

Results for "Theremin"

- Initial Rate of Success
  4/27 (successes in attempts) = **14.8%**

- Prevent Dictation of "Sarandon" for Theremin
  15/58 = **25.8%**

- Add "Theremin" Word to Speech Dictionary
  134/165 = **81.2%**

- Add "Theremin" to Loaded Grammar (so that dictation (wildcard) speech recognition isn't used
  97/101 = 96%

Results for "Theremin"
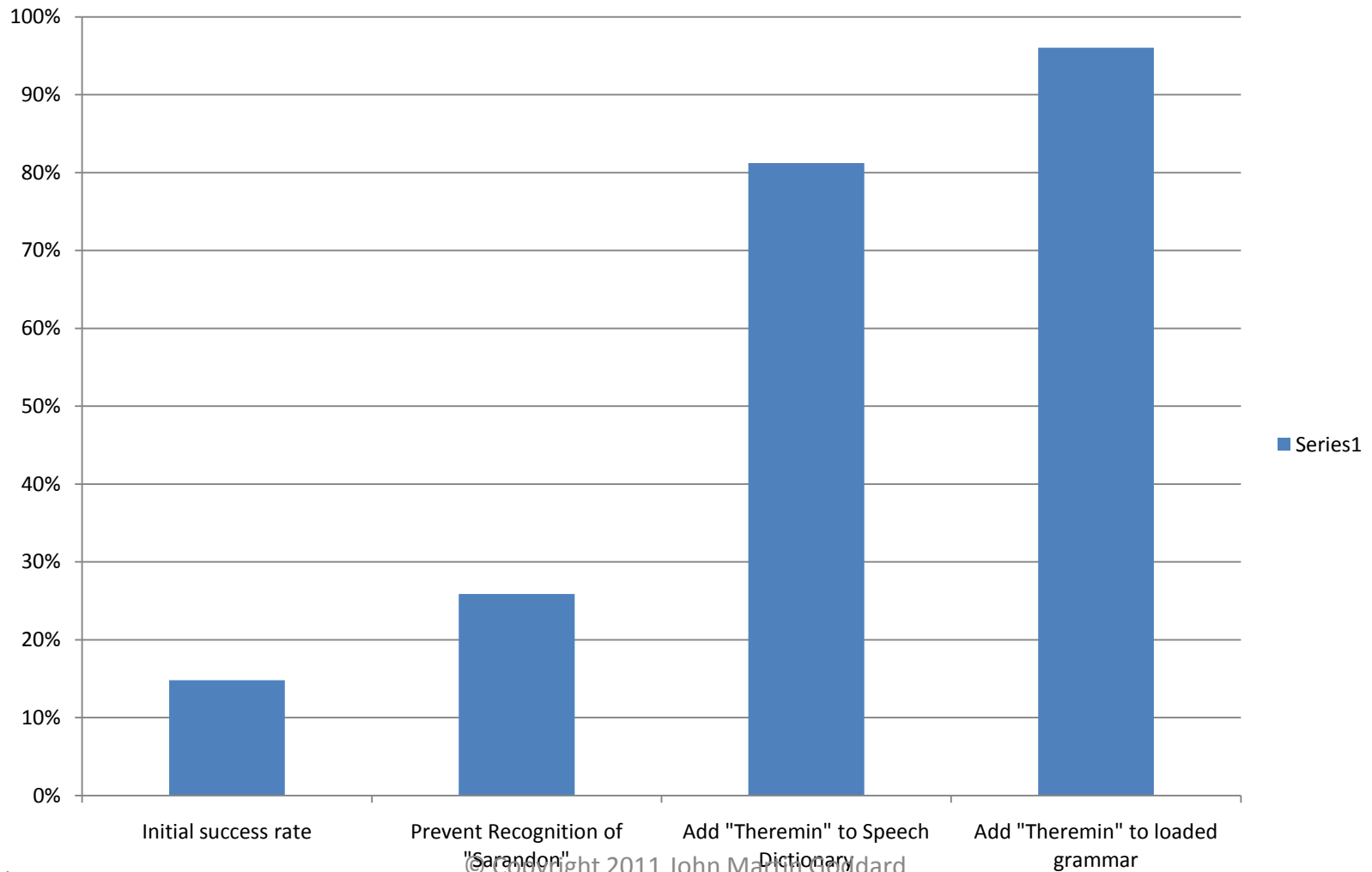
"Theremin" Experiment revisited, to show that improvements were not learned

- Add "Theremin" Word to Speech Dictionary

  AND

  Prevent Dictation of "Sarandon" for Theremin
  - Right from start.
  - 74/80 (successes in attempts) 75/80 = **93.75%**

- Add "Theremin" to Loaded Grammar (so dictation (wildcard) speech recognition isn't used)
  - Improved to: 148/153 = **96.7%**

Interpretation

- Using dictation speech recognition for names already assigned, does perform reasonably well (>90%)

- Adding names to the loaded grammar works better than without having them loaded (and, hence, relying on dictation recognition)

- So, why not *always* load the names into the grammar?

Interpretation, cont.

- Why not load new names into the grammar every time?

- The names are not always correct.

- The user should choose when to get all the names in the session and load them into the grammar.

Experiments with Homophones

- Pair of words that sound the same …
- … but are Spelled Differently
- Examples: Pair and Pear
- Selected from over 1000 homophone pairs only those having to do with music recording
- Determine which is the default
- Then try to with elaboration to assign the other spelling
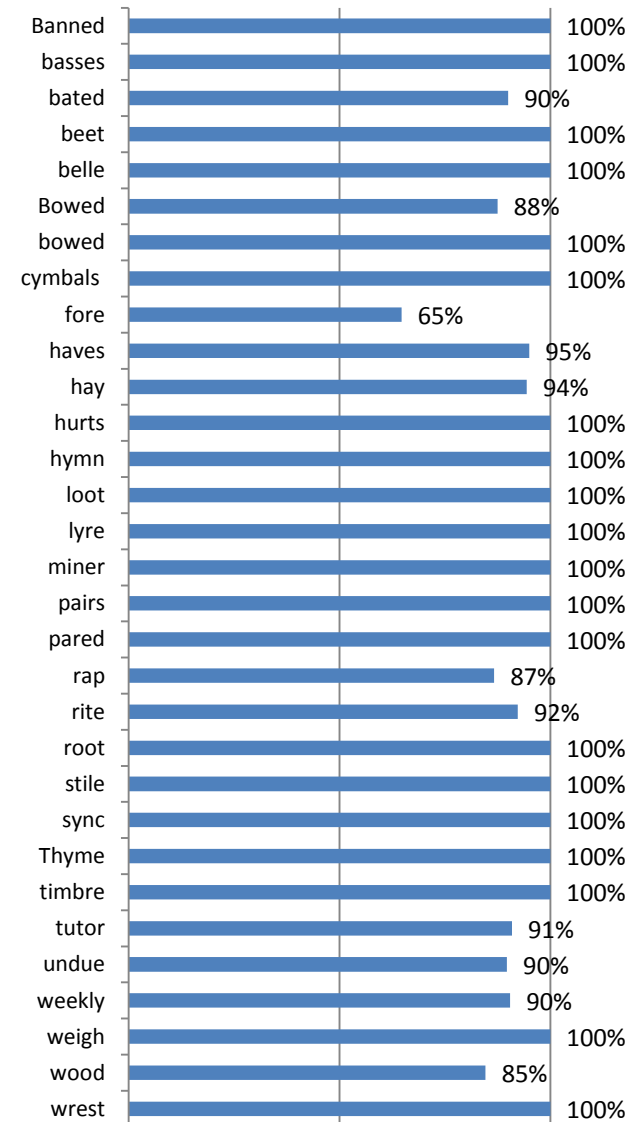
elaboration success rates
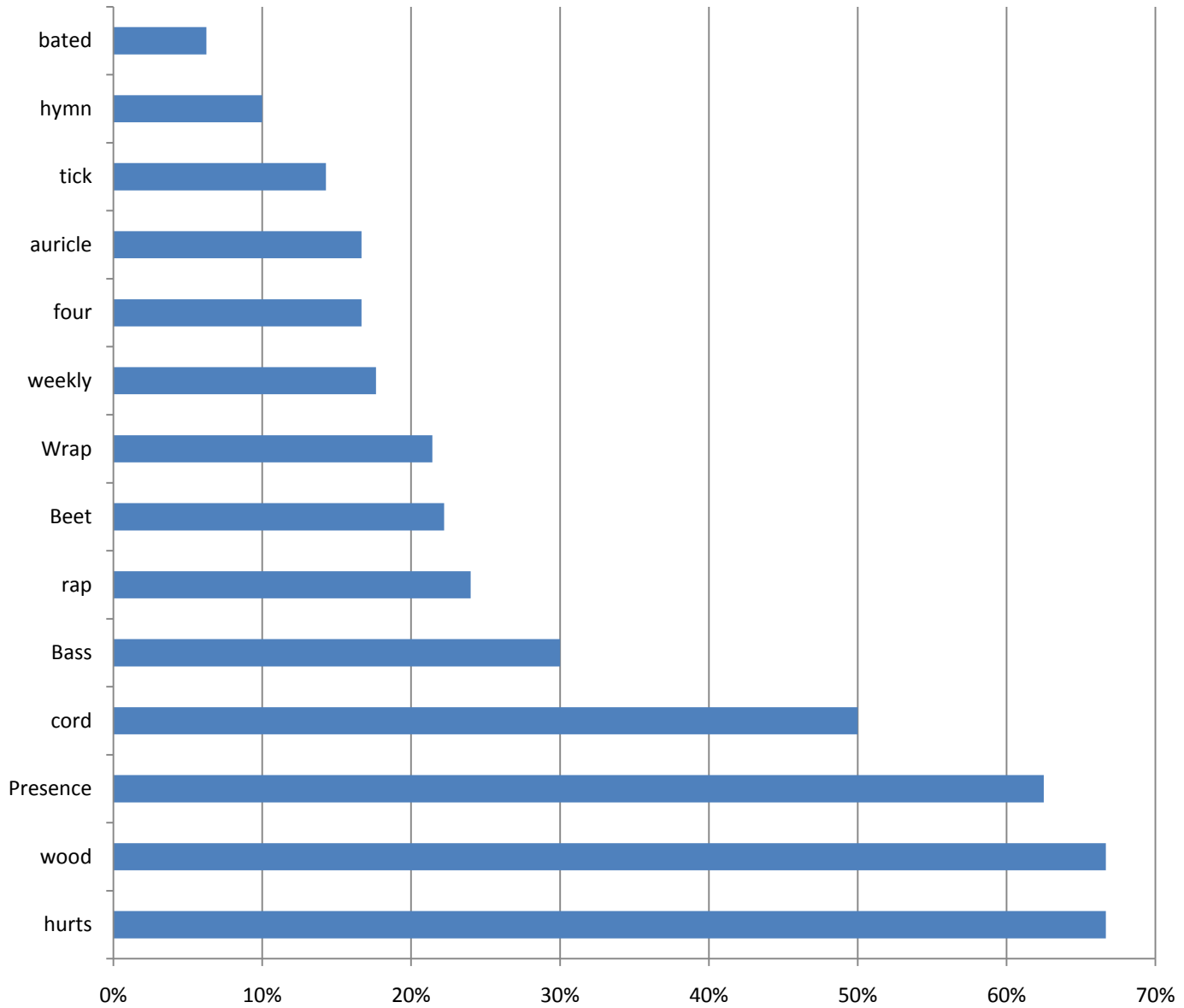
Interpretation of Elaboration Results

- Elaboration works sometimes
- There seems to be a skew toward assigning Proper Names to named entities, based on the entire naming command phrase
  - Wei instead of way
  - Ryan or Orion, instead of rhyme
  - Sarandon, Shanti, Peres, (think of others)

After Loading into Grammar

- The non-default spelling of the Homophone pair was recognized with high accuracy

Recognition Accuracy
(after loaded into Grammar)

Word	Accuracy
Banned	100%
basses	100%
bated	90%
beet	100%
belle	100%
Bowed	88%
bowed	100%
cymbals	100%
fore	65%
haves	95%
hay	94%
hurts	100%
hymn	100%
loot	100%
lyre	100%
miner	100%
pairs	100%
pared	100%
rap	87%
rite	92%
root	100%
stile	100%
sync	100%
Thyme	100%
timbre	100%
tutor	91%
undue	90%
weekly	90%
weigh	100%
wood	85%
wrest	100%

Interpretation of Results of Loading into Grammar

- Once a name is correctly assigned (by whatever means), it can be loaded into the Speech Recognition Grammar for accurate recognition