Pardon Me, Your Computer's Showing:
Using speech to speed and streamline desktop computing

Before I get started I'd like to ask a couple of questions.

How many people here regularly use speech controls for cell phones, smart phones or PDAs?

How many use speech on a desktop or laptop computer for dictation?

For controlling a computer?

How many expect to use speech to control a desktop or laptop computer this year?

I'm here to talk about how to use speech to control a computer in a way that's natural and efficient.

First I'm going to show you how I use speech. I'll do some word processing, make a table and graph, move windows around, open dialog boxes, and access the Internet.

The speech interface software I'm using is Redstart Systems' Utter Command. Some of you might have seen some features of Utter Command at a SpeechTEK talk in New York last summer. We're now just finishing up beta testing. This version uses the NaturallySpeaking speech engine. We replace all the NatSpeak commands.

I'm going to move along quickly here -- I'll explain more later.

Speech On Word Open
Today Date Long
Another Graph
we saw many many little birds skimming over the water
4 Befores
swooping over the marsh
2 Before 1 Befores Delete
around
3rd Word 2 Words Bold
Go Home
This morning
Right Delete w
Another Graph
This afternoon we assessed productivity Colon
**Another Graph**
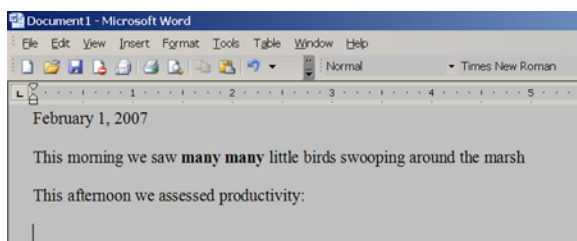
Under t w Close

Excel Open Max
Cell Charlie 6
Days Short Tab
Cell Bravo 7
Cap Brainstorms

Tab 1 Tab 2 Tab Short
4 Point 7 Tab Short
6 Point 5 Tab Short
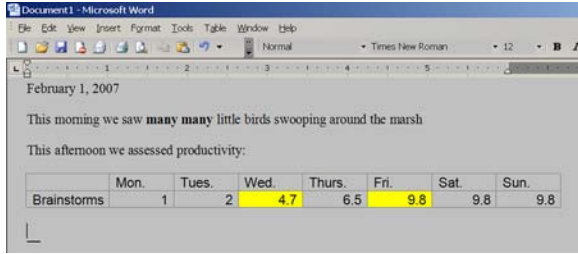**9 Point 8 Tab Short Repeat 3**
Alternate Equals Enter

8 by 91 No Touch
Cell Echo 7 Touch
Cell Golf 7 Touch
5 2 Letters
1 Ups

Function 11
Control Left (corners selected)
Shift Control Papa 1 5 Enter
This Bold
Control Right (bars selected)
Control 1 Under Delta
3 Control Tab
Under Victor Enter
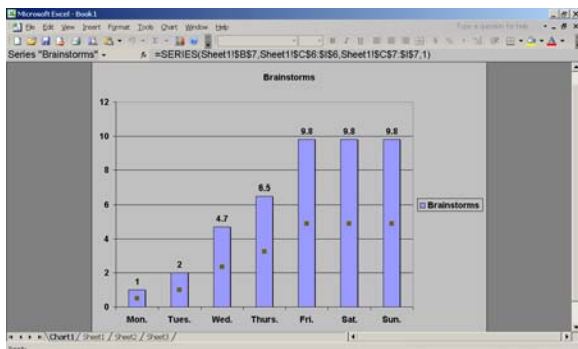
Control Page Down
Copy to Word

Size 90 by 40
Window 0 by 50
Window 0 by 0
Screen Clear
Word Max

Excel Close No
Word Close No

Control Panel Delta Down Enter
New York Times Site
Screen 3
Task Processes

Window Close Times 4 Speech Off

(This demo required 46 commands. Doing this demo with a keyboard and mouse would require 116 commands.)

You might notice that our syntax is different from existing speech interfaces and we tend to get more done with each command.

Another thing that makes us different is our commands are not program specific – these commands work across all programs. As you saw, these generic commands allow you to do what you want even in a complicated program like Excel.

Keys to speeding and streamlining computer control

There are three keys to using speech commands to speed and streamline computer control: minimizing steps, making commands easy to remember, and enabling combinations.

SLIDE 2
Keys to using speech to speed and streamline computer control

Utter Command is strong in all three areas because it is anchored by an efficient grammar – Human-Machine Grammar -- that follows the way the brain works and is designed for the purpose of controlling a computer.

SLIDE 3
Human-Machine Grammar

3

Let's take a detailed look at each of the three keys to speeding and streamlining computer control.

Minimizing steps

Controlling a computer using speech has, not surprisingly, closely mimicked the input methods we're familiar with – the keyboard and mouse. This has made for a disappointing decade and a half of speech control of the desktop.

Although it's good to tap existing knowledge, it's also important not to let experience confine new methods of communication.

It takes 11 keyboard and mouse steps to paste into a Word document a picture that's buried four levels deep within the file system.

The steps are necessary because the keyboard and mouse use restricted resources – a limited number of key combinations and a limited amount of screen space to click on.

It doesn't seem slow going through a series of menus because it's familiar, and things are busily moving along. But when you think outside the silicon box it becomes apparent that there really shouldn't be as many steps involved.

Instructing a person using as many steps as it takes to retrieve a picture using the mouse and keyboard might go something like this: Stand up, walk to the file cabinet, open the third drawer, go to the "B" section, find the bird folder, open the bird folder, find the picture subfolder, find the Redstart Singing picture, take it out of the folder, put it in my hand, let go.

Instructing a person is usually much more efficient: can you get the Redstart Singing picture? It's obvious that it's inefficient to instruct a person step-by step. It's inefficient to instruct the computer step-by-step as well. But it's not as obvious because inside the box these steps, required by the GUI, are comfortably familiar.

4

Most of today's desktop speech software follows in the footsteps of the keyboard and mouse, pausing between steps. But it doesn't have to be that way.

Speech On Word Open
Under i p f
Bird Pictures Folder
1 Right Enter
Window Close No 3 Seconds Break

Here's an even faster way that doesn't use any key commands specific to Word:

Redstart Singing File
All Copy to Word New
Window Close No Window Close Speech Off
(UC 8, key/mouse 28)

Eleven steps versus two steps is a 550% difference in efficiency.

SLIDE 9
Retrieving a Picture – Efficient Commands

Let's look at another task – sending email. Addressing an email to two people, CCing a third and writing a subject and first line takes 12 keyboard and mouse steps.

SLIDE 10
Address email and add subject, first line

It can be dramatically fewer using speech.

SLIDE 11
Address email and add subject, first line

Speech On
Express David Linda CC Kim
Paris trip
1 Tab

Window Close No Window Close Speech Off
(UC 5, key/mouse 15)

Road versus Air

The potential differences between speech input and keyboard and mouse input are akin to the differences between road and air travel.

Following a road via airplane is faster than driving, but the real power of air travel is the ability to fly point-to-point and travel any route.

Using speech to follow in the footsteps of the mouse is akin to using an airplane on the ground. It's accommodating the GUI rather than moving beyond it.

Remembering Commands

As we've seen, steps are very important – the speech interface is not easy to use unless it's fairly efficient. But there's more involved in making the speech interface easy to use than keeping steps to a minimum. Those steps have to be easy to remember.

I'm going to talk about talking for a minute, then we'll come back around to remembering commands.

When discussing philosophy you need a lot of words on tap. You're exploring a complicated subject and changing gears on-the-fly based on agreement, argument, and different types of understanding. Vocabulary matters. Subtle timing matters. Dramatic gestures matter.

And they all take cognitive energy. This is okay when you're arguing philosophy because arguing philosophy is what you're focused on.

Talking to a computer is a means to get something else done. Here cognitive load matters. You don't want to use half your brain power remembering commands when you're trying to concentrate on writing a talk or doing research or putting numbers in a spreadsheet or editing a picture or even organizing your email.

History

Remembering speech commands is a long-standing problem.

In the late '90's NatSpeak architect Joel Gould attended a Boston Voice Users Group meeting and posed the question "How do you remember commands?"

A year later, Joel illustrated the problem during a demo of a new version of NatSpeak. He changed font sizes and colors. The audience, which included regular and long-time NatSpeak users, oo'ed and ah'ed. Joel asked if these were commands we'd use, and many of us emphatically said yes, these are great, we're going to use them all the time. Than he said "you do have these – these are in the old version. Nobody uses them".

A few years later, I talked to someone who had done NatSpeak trade show demos and she talked about how hard it was to memorize the long Excel commands that looked impressive.

When speech trainers get together a similar topic comes up again and again: why do people so often ask trainers to be supplemental memory, to simply feed commands as needed?

All of these anecdotes point to the same problem—the type of speech command language we're using imposes too heavy a cognitive load.

What speech command language needs to accomplish

The language that's required to a control a computer isn't nearly as complicated as the vocabulary needed to talk about philosophy. It's not even as complicated as what you would use for idle chat with a friend.

Using speech on a computer is more like giving orders in a fast food kitchen -- "Two Fry",

calling a play on a football field -- "Counter Trey Right",

a police dispatch -- "Unit 26, Code 11-31, 13th and Vine",

or an air traffic control communication -- "Delta 265, clear to land, runway three zero".

Probably the most comparable language to the ideal for controlling the objects on a computer screen are the commands fictional magical characters like witches and wizards give to objects around them: "Chair Dance", "Teeth Grow".

Here's a similar exercise: picture me as your stereo controller – that's all I do – what would we work out as a speech interface? Would you say "Can you please change the channel to my second favorite blues station"
Or "Blues 2"?

The language humans work out in command-and-control situations is a natural for issuing the same commands over and over to a computer.

This type of language – efficiently worded instruction -- is what the Human-Machine Grammar underlying Utter Command is all about.

Human-Machine Grammar

Human-Machine Grammar and Utter Command were informed by cutting edge research in cognition, linguistics, networks, and interfaces. Important factors include how the brain processes words, the memory phenomenon of chunking, and the network phenomenon known as six degrees of separation.

Human-Machine Grammar has no synonyms, uses logical rules to minimize wording possibilities, and follows the way the brain uses language.

In Practice

Utter Command has 257 command words. Ninety-seven of these identify keyboard keys, leaving 160 words to learn to master all of Utter Command. These words are by design easy to remember, and commands are consistently constructed according to 16 grammar rules.

8

Learn about a third of the words, many of which are common and obvious, and a handful of general rules, and you'll find yourself humming along nicely saying things like 5 Down, Line Cut, and Window Close.

Here are the 60 most commonly used words of the Utter Command vocabulary.

Most are very familiar -- copy, window, undo -- and many are paired -- up, down, left, right, before, after, top, bottom, open, close.

And here are the most common Human-Machine Grammar rules:

- Match words used for a command as closely as possible with what the command does. For example, "Window Close".

- Be consistent – always use the same words for the same actions – "Line Bold", "Line Copy".

- Eliminate synonyms – the smaller the command vocabulary the easier it is to remember and use.

- Follow the way people naturally adjust language to fit a situation – commands that follow the way the brain works are easier to remember and use.

- And follow the order of events as they'll be carried out – generally you identify or select an object like a line or paragraph, then carry out an action – "Line Delete". This dramatically cuts wording possibilities and is easier because it follows the way you think about doing the task.

Put the words together with the rules and you get commands like "Speech On", "Line Copy", "3 Before", "Window Close", "Screen Clear".

Combining commands

This brings us to the third key to using speech to speed and streamline computer control: enabling combined commands.

Utter Command's terse grammar makes it possible to say several commands in a single phrase, which speeds computing considerably. You've heard me say combined commands like "3 Befores Delete" and "Word Close No Speech Off".

The results of a good grammar

I'd like to stress one other thing. Like the keyboard and mouse – Utter Command works across all programs. It allows you to do everything you do with the keyboard and mouse, usually faster. The grammar generally cuts the number of steps by 100 to 200 percent and in some cases many times that.

Here are some typical commands – you've seen me do all three of these.

SLIDE 33
Computer Command Steps

Take a look at this first example – it takes a single UC command to bold words before and after the cursor, but it's three steps using the keyboard and mouse and four steps using speech that follows in the footsteps of the keyboard and mouse.

And as you've seen, the difference in steps is more dramatic with tasks like preparing an email message

SLIDE 34
Computer Command Steps

and starting a program and opening a file that lives four folders deep in the file system.

SLIDE 35
Computer Command Steps

We also know that 74% of users prefer a structured rather than natural language approach to speech recognition -- that's from a Carnegie Mellon study.

SLIDE 36
Carnegie Mellon study

Taking flight

So, to sum things up, there's a lot of good news here. Using a command language that's based on efficient human process rather than the constraints of the keyboard and mouse allows you to take only as many steps as necessary. If you don't have to think between steps, there's no need for separate steps.

It makes command easy to remember, which frees your brain to do tasks on a computer. And it enables combinations – which further speeds things up.

10

Few steps, easy-to-remember commands, and combinations allow you to fly rather than take the ground route.

Beyond the keyboard and mouse

And here's where things get interesting.

The real power of speech is the ability to command the computer in ways not possible using the keyboard and mouse alone. Getting away from the constraints of the keyboard and mouse – once you recognize you can be free from them – gives you new abilities.

I'll demonstrate four types of commands that go beyond the keyboard and mouse in some way: speech links, document access and navigation, favorites-like lists, and a virtual calculator.

Speech Links

Augmented reality is inherent in speech recognition. Here's a simple example.

Speech On
Demo 9 File
8 Seconds Break

Here's a Web address that's not linked, so you can't click on it. But if you're using speech you don't need the link – it's inherent in the command.

Line Copy to Firefox
9 Seconds Break

This makes any Web address a speech link whether or not it's linked.

Here's another type of speech link – searching for any word on dictionary.com

Word Open
2 Down 1 Afters
This Dictionary Search

Window Close Times 2 Speech Off
(UC 9, key/mouse 25)

Documentation

This principle can also speed getting around documentation.

11

Finding what you want in documentation is a classic problem. We think one reason is the number of steps involved in finding what you want. Finding instructions to add a special character to your Word document using the standard electronic documentation, for instance, takes 8 steps if you go through the table of contents and 7 steps if you use the search function.

SLIDE 41
Word Help – Insert a Special Character

The Utter Command manual table of contents -- printed or electronic -- provides all the information you need to use a single speech command to go directly to any section of the documentation.

Speech On
UC Full 10 Point 9
7 Seconds Break

This is the Full Reference section -- say you want a more detailed explanation of the same commands from the Lesson section. It's one step to get there.

UC Lesson 10 Point 9
2 Seconds Break

And here's the same section in a quick reference cheat sheet.

UC Quick 10 Point 9
Window Close Speech Off
(UC 7, key/mouse 18)

SLIDE 42
Beyond the Keyboard and Mouse

List Commands

Utter Command contains a utility that allows you to keep favorites-like lists of Web sites, files and folders you can access in a single command.

Speech On
Slash Dot Site
Demo 1 File
Windows Bird Pictures Folder
Word Bird Pictures Folder
Window Close Times 4 9 Seconds Break

UC List also contains a list of keywords that works with the Find function of any program. You can arrange this list as you like, deleting and adding anything you want.

Demo 8 File
Find Scene 2

Find Scene 3
Find Courtney
6 Seconds Break

These commands work because "Scene 2", "Scene 3" and "Courtney" are all on my keyword list.

Window Close Speech Off
(UC 11, key/mouse 41)

Virtual Calculator

All the functionality you've seen until now is in the first version of Utter Command. This next demonstration is just a proof-of-concept prototype.

It's a classic example of the limits of the GUI that, even though Windows comes with a calculator, people still buy separate calculators. It takes fewer steps to reach for a physical calculator than to click more than once to get to the calculator applet.

Using speech you can change the equation. You don't have to reach for the virtual calculator at all, and it presents the answer exactly where you want it in your choice of ways -- the work and the answer, or just the answer.

Speech On Word Open
Five Times Seven Five Three Equals
Another Line
Five times Seven Five Three Answer
Another Line 2 Seconds Break

Or you can have it both ways.

Five Four Divided by One Point Two Wait Answer
**Another Graph 2 Seconds Break**

You can also keep going.

5 Times 5 Equals
Times 1 6 7 Equals
Times 9 6 5 Equals
3 Enter 2 Seconds Break

This should sum things up.

Utter Command has 256 words
Comma 97 of which are key names Comma making New Line
2 5 6 Minus 9 7 Wait Answer
words
to learn to master all of Utter Command Period
Another Graph
A third of the words Open Paren

1 5 9 Divided By 3 Wait Answer
Close Paren will get you up and running in no time Period

Document1 - Microsoft Word

File Edit View Insert Format Tools Table Window Help

Normal   Times New Roman

5 * 753 = 3765
3765
45

4 * 4 = 16 * 165 = 2640 * 965 = 2547600

Human-Machine Grammar has 256 words, 97 of which are key names, making 159 commands to learn to master all of Utter Command.

A third of the words (53) will get you up and running in no time.
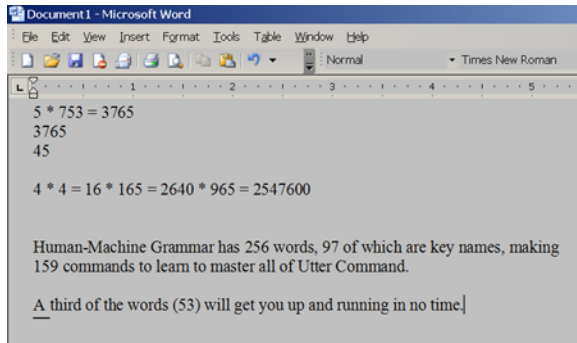
Window Close No Speech Off
(UC 14, key/mouse 95)

For more information, feel free to contact me at kim@redstartsystems.com.

SLIDE 43
For More information

The Human-Machine Grammar rules and dictionary are posted at the Redstart Systems site. There are also talks posted on the site that go into more detail about the thinking and science behind Human-Machine Grammar. And there's also advice for software makers who want to make sure they are ready for speech input. We're encouraging everyone who writes speech commands to use Human-Machine Grammar.

When you go to our site you'll find a mature structured grammar that was developed over the past decade in a real-world environment.

You heard me say 101 speech commands during this talk. If I'd used the mouse and keyboard instead, I'd have had to use 348 commands. The speech demo portions of this talk takes about 8 minutes. It takes about 17 minutes to use the keyboard and mouse for those demos -- just over double the time -- that's the productivity gain.

What you've seen here today is the tip of an iceberg. Human-Machine Grammar and Utter Command are baseline technologies -- change agents -- that will enable much more.