It's All about the Interface:
Speech Recognition That Works for Both Human and Computer

Massachusetts Institute of Technology (MIT) Independent Activities Period (IAP) seminar

January 17, 2006

Kimberly Patch
President, Redstart Systems

Note: This talk includes several short demonstrations of the Utter Command speech interface. These commands appear in bold.

Greetings. I'm here to talk about the speech recognition interface and how it can solve some classic computer interface problems.

I've been a journalist for many years, and I've also used speech recognition hands-free for many years. I started using speech recognition 12 years ago because of repetitive strain injuries in my hands. For a long time I was frustrated by the speech Interface. The commands were sometimes long, there were multiple commands for the same function, and there was no apparent pattern to the way commands were structured — all of which made it hard to remember the commands.

As a writer I spend a lot of time thinking about words. In the past few years I've been able to make the speech interface much less frustrating by thinking about the words — the speech human-computer interface.

I'm going to cover four topics during this talk:

SLIDE 2

I'll talk about
- the key to using speech as an interface to control the computer
- the classic computer interface problems a speech interface can solve
- the rules and words, or Human-Machine Grammar — that underlie an effective speech interface
- and Redstart Systems Utter Command speech interface software

Throughout the talk you'll see some short demonstrations of Utter Command, and at the end we'll give you a longer demonstration of sending an email about a trip to Rome. The talk and demo will take a little under an hour, leaving plenty of time for questions.

First, however, I'm going to read to you — very briefly — from Harry Potter and the Goblet of Fire by J. K. Rowling.

Archrivals Harry Potter and Draco Malfoy are facing off in a Hogwarts hallway, wands drawn, trusty friends Hermione and Goyle close by:

> *For a split second, they looked into each other's eyes, then, at exactly the same time, both acted.*
>
> SLIDE 3
>
> *"Furnunculus!," Harry yelled.*
>
> SLIDE 4
>
> *"Densaugeo!" screamed Malfoy.*
>
> *Jets of light shot from both wands, hit... in midair, and ricocheted off at angles — Harry's hit Goyle in the face, and Malfoy's hit Hermione. Goyle bellowed and put his hands to his nose, where great ugly boils were springing up — Hermione, whimpering in a panic, was clutching her mouth... [Her] front teeth... were... growing at an alarming rate*

Now let's get back to the real world, where words aren't usually so powerful and speech recognition technology has not yet lived up to its considerable potential for speeding and simplifying computer use.

Today's speech engines are relatively quick and precise in recognizing dictated words. The speech interface, however — the command words used to control a computer —

SLIDE 5

to open programs
close files
move windows
cut and paste
bold words
delete paragraphs
click menu items
turn up the volume
access web pages
or select text

— is a lot more taxing than it should be. Remembering commands is one problem. Juggling spoken commands and whatever you're using your brain to do on a computer is also taxing.

The solution to both these problems — and the key to unlocking the potential of speech recognition — is deceptively simple.

The key to using speech to control a computer is not to talk to the computer.

SLIDE 6

Don't talk to the computer.

The reason is the world of the computer is different from the real world. In the world of the computer, objects like programs, files, windows, the cursor, words, lines of text, and menus — can hear you.

So, the key to using speech to control a computer is not to talk to the computer. Instead, talk to the objects on the computer screen.

SLIDE 7

Don't talk to the computer. Talk to the objects.

When you use speech to close a window in the real world you have to get a third party involved — I might ask Bill, "Bill, will you close the window?" — or to be succinct and bossy, just say "close window".

In the world of the computer you can skip the middleman and talk directly to the objects: "Window Close".

Bill is now going to go through a series of commands — notice how he's talking to the objects.

Wordpad Open
Wordpad Close
Word Open

Help Folder
Window Close
Window Close
Notepad Open
Window Maximize
Window Restore
seven cats in a row
Line Duplicate Times 7
Line delete
4 Down
3 Lines
Window Close No

There is a big advantage to being able to talk directly to objects.

SLIDE 8

Saying exactly what you want to happen — "Window Close", "3 Lines", or "Line 3 Back"— is cognitively easier than coordinating an action between an object and a third party — "Close Window", "Select next 3 lines", or "Move line back three lines".

It's natural to make things happen directly when you can.

It may not seem natural to talk to the objects on your computer screen, but that's just because we lack experience — objects in the real world generally can't hear you. The average window, door, tree, garbage can, orange, breadbox, coffee cup — none of them respond when addressed directly.

Although most of us lack experience in addressing objects directly, there is precedent that shows that when objects can hear — we do naturally address them directly.

Imagine, for instance, what Harry Potter would say if he wanted to use magic to close a window or make a chair dance — something along the lines of close or dance if he were pointing to the object with his wand or "Window Close", or "Chair Dance" if he had to identify the object using words as well. If Harry tried to get a window to close or a chair to dance using spells along the lines of "Close Window" or "Make Chair Dance" — you'd wonder who he's talking to.

The key to making the speech interface easy to use is recognizing that the on-screen world harbors an element of magic. Every object you see on the screen — windows, programs, cells, paragraphs, buttons and slider bars — has the ability to listen and act.

Let's look a little more carefully at the spells from the earlier reading.

SLIDE 9

Furnunculus causes boils to break out
Densaugeo causes rapid tooth growth

Harry Potter and Draco Malfoy were using their wands as pointing devices to identify the objects they wanted to act on — each other. The actions they want to carry out — boil growth and rapid growth — are spoken.

In Malfoy's case, however, he also had to speak to identify an object finer than just the person he was pointing at — his spell is made up of two Latin roots. The first — dens, or tooth — refines the object from the entire person he is pointing at to just teeth, while the second indicates the action — rapid growth.

SLIDE 10

So, magic and easily spoken computer commands are quite similar: object, action

This object, action word order is cognitively easy because it's the way you think. You first think of the object — window, then what you want to do to it — close. When you switch the command around you have to hold the object in memory while you say the action, then retrieve the object.

There's something else to notice about the way magic and computer commands can be worded.
Although objects in magical worlds and on the screen can hear you, they don't have feelings.

SLIDE 11

This makes it quite OK to order them around using succinct object, action commands like "Window Close", "Go Top" or "Furnunculus" rather than the more polite, and wordy: "will you close the window", "go to the top of the document", or "please make ugly boils appear".

There's a distinct advantage to succinct, bossy commands.

SLIDE 12

Fewer words are easier to remember, say, and cultivate as habit.

There's another advantage to talking to objects when you're talking to a computer.

When you're talking to an object it's fairly obvious that you're talking to an object.

This type of command — destined for the computer — is less likely to be misinterpreted by a person.

SLIDE 13

In Harry Potter novels people don't get mixed up about whether you are telling someone to get your broom or commanding the broom to fly into to your hand.

Say Close Window in a crowded room and people have to decide whether to act. Say Window Close in a crowded room and if people are used to magic — or speech recognition — they'll expect the window to close itself.

To sum things up, the keys to using speech to control a computer are talk to the objects, and use direct, succinct language in doing so.

SLIDE 14

So now that we've got a nice succinct vocabulary that is easy to remember and comfortable to use, speech recognition can live up to its potential.

There are some classic interface challenges that have bedeviled users for quite some time that this type of speech control addresses.

SLIDE 15

I'll talk about four of them: folder and file access, command consistency across programs, keeping command steps to a minimum, and fluid switching among programs

I'll show how speech recognition addresses these challenges, and in doing so, speeds and simplifies computer use.

We'll be demonstrating this using Utter Command software — this isn't quite ready for prime time yet — it's scheduled to be out this summer.

First, there's the file system. If you have 100 files how many clicks do you need to open the average file? How about if you have 1000 files? How about if you have 10,000 files? How about on a network that hosts 100,000 files? Computer file systems don't scale very well.

Utter Command includes a utility that allows you to open any folder or file using a single speech command.

This allows you to go directly to any folder from whatever program you are in.

We'll call up a folder named "Demo".

Windows New

Here's what happens when we call up the Demo folder from Windows Explorer.

Demo Folder to thank

And here's what the same command does when you're in WordPad. It calls up WordPad's open dialog box to the folder you specify.

Wordpad Open
Demo Folder
Close
Window Close

This is similar in all programs with open file dialog boxes. Here's Excel.

Excel Open
Demo Folder

Help Folder
Close

You can also call up files directly.

Window Close
Window Close

The file you call up will open in its default program.

Demo 1 File
Demo 5 File

Window Close No
Window Close

Second, different programs often do the same things differently. I need a show of hands here. Anybody know what you have to click to paste the date in Word? How about Excel? How about Wordpad? OK, how about keyboard shortcuts for dates in Word, Excel and Wordpad?

SLIDE 16

The keystrokes and mouse clicks for pasting the date in these programs are inconsistent and often require several steps. The speech command remains one consistent command in all programs.

The computer keyboard has a real estate problem. Keys are in short supply. The computer mouse has a real estate problem. The mouse requires something to click on and clickable items take up screen space, which is also in short supply. The speech interface has no real estate problem. There are plenty of words and phrases available. This enables commands to more easily be used across programs — like this.

Wordpad Open
Date Short
Excel Open
Date Short
Word Open
Date Short

Third, computers require many steps. To close these windows using a mouse or keyboard I have to tell the computer to close the window, then, in a separate step, tell the computer whether or not I want to save the file.

Speech commands, however, can be combined. Closing a file and saving it or not saving it in one step rather than two cuts the number of commands in half.

Word Close No

Closing a file that's behind another file and saving it or not saving it in one step rather than three is two third fewer commands.

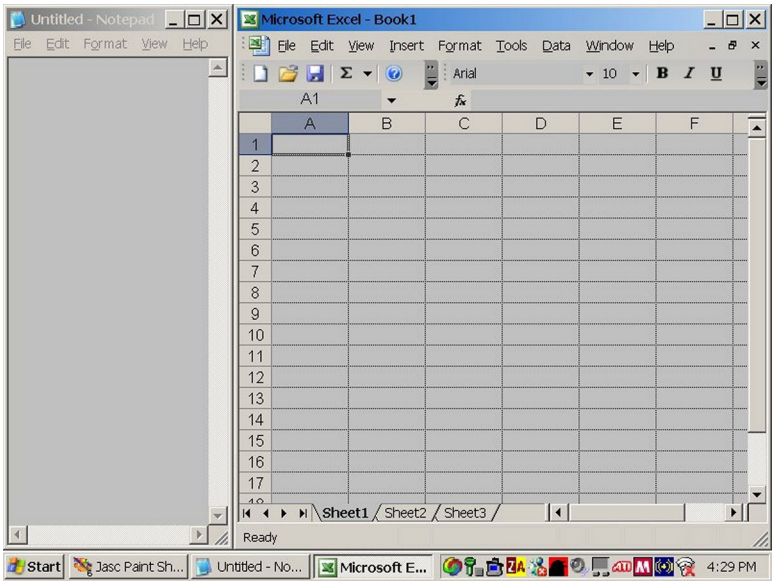
Wordpad Close No

And sizing a pair of windows precisely requires a fair amount of mousing.

These commands allow you to size two Windows at once, specifying exactly where you want the windows to meet on the x or y axis.

Notepad Open

Window 2 3 X 30



Window 2 3 Y 50
Window 2 3 X 70

Window 2
Window Close No

This command allows you to size a window.

Size 50 by 50

And this one to move the window so it's top corner is at the top of the screen.

Window 0 by 0

The fourth classic challenge is fluid switching among programs. With speech you can often get anywhere from wherever you are using a single command.

You can be using Excel, you can be using WordPad, you can be using Notepad, or you can be using no program at all. You can turn up the volume.

Volume 40
Volume 90

You can set a timer. (1... 2...)

2 Seconds Time

You can set a reminder. (1... 2... 3...)

3 Seconds Bother Boss

You can send mail.

Eudora Karen

Document Close No

You can access the Internet.

Google Site
Slash Dot Site

And as a bonus, you can combine some of these advantages. Here's a combined command that works from anywhere.

Eudora Kim Lisa CC Eric
Document Close

Window Close No
Window Close No
Window Close

So, to sum things up

SLIDE 17

The speech interface can improve folder and file access, it can improve command consistency across programs, it can minimize command steps, and it can enable fluid switching among programs.

Now I'm going to talk a little about Human Machine Grammar — the system of words and rules that underpins Utter Command.

SLIDE 18

How many people have read the talk titled Utter Command that was referred to in the notice to this talk?

This next section is a recap of the portion of that talk that described Human-Machine Grammar. I'll go over a couple of important points, talk about the books that informed Human-Machine Grammar, and briefly describe Human-Machine Grammar.

SLIDE 19

In addition to allowing you to talk to objects, a good speech interface must contain commands that are easy to learn and remember, so you can reserve as much of your brain power as possible for the tasks you are using the computer to do.

A speech interface designed for communications between a person and a computer is also necessarily different from the speech interface we use among people. People can adapt to each other and to the task at hand; time-saving jargon evolves naturally in human-human communications. Computers, however, don't adapt.

This makes it critical, in human-computer communications, to work out a command language ahead of time.

SLIDE 20

Human-machine grammar is made up of a relatively succinct dictionary of words that are used to build commands according to a concise set of grammar rules.

The system is relatively easy for humans to learn and remember. It also allows computers to respond to the commands without having to decode natural language or be loaded down with very large sets of synonymous commands.

This grammar system has been informed by four fairly different books:

SLIDE 21

Words and Rules, by Stephen Pinker
Linked, by Albert-Laszlo Barabasi
The Psychology of Everyday Things, by Donald Norman
and *The Humane Interface*, by Jef Raskin

SLIDE 22

Words and Rules points out how highly ordered language is, how difficult it should be to learn, and how natural it is for humans to learn.

In working out a speech computer interface it's important to go with the general flow of human language to make things seem easy. Go against that flow, and commands become much more difficult to remember and use.

SLIDE 23
Linked delves into the structure of the many networks in the world — including the Internet, social networks, and relationships among words in a vocabulary.

The key piece of information from this book is the more connected a network, the easier it is to get from one point to another — the six degrees of separation concept.

This translates to the command vocabulary used by human-machine grammar as well. The smaller and more connected a vocabulary, the easier it is to think of the words and structure you need for a given command.

SLIDE 24

The Psychology of Everyday Things contains many insights about the interfaces all around us.

It points out how important visual cues are, and it hammers home the point that when people repeatedly make mistakes that seem stupid — like pushing the wrong side of a door or turning on the wrong stove burner — the root cause is almost always a design flaw.

The same is true of speech. A speech interface that is difficult to use points to a design flaw.

SLIDE 25

The Humane Interface contains many insights about computer interfaces.

It stresses how easy it is to trip someone up by messing with habits, and points out the weaknesses of today's graphical user interface. It's ridiculous to require a different series of clicks or a different key combination to carry out the same function — like adding the date — depending on which program you happen to be in.

This inconsistency means that once something becomes habit you'll start making mistakes like pressing the key combination to paste the date in Word when you are in WordPad, precisely because you're not thinking about it.

The great thing about the speech interface is that it can address some of the inherent inadequacies of the graphical user interface — if it is designed to do so.

SLIDE 26

The important elements of command language are words, context, and word order. Human-machine grammar leverages all three.

Here are the basic human-machine command building guidelines:

SLIDE 27 — Command-Building Guidelines

Match commands to meaning
Use words the user sees on the screen when possible
Balance ease of saying and ease of remembering
Keep commands succinct
Don't use synonyms
Conservative words by using multiple meanings

We match the words used for a command as closely as possible with what the command does, and we use words the user sees on the screen when we can. This makes commands easier to remember.

We keep in mind that the ease of saying a command is always important, but becomes even more important the more often a command is used. In contrast, the ease of remembering a command is always important, but becomes even more important for commands that are not frequently used.

We use one-word commands very sparingly because they are apt to get mixed up with text. Beyond one word, however, we keep the number of words used in any given command to a minimum, because this makes them easier to remember, say and combine.

One tenet of human-machine grammar is there are no synonyms. This makes the vocabulary smaller, makes it easier to remember and predict commands, and makes it possible to combine commands, which speeds everything up.

Another important way to keep vocabulary to a minimum is not using a new word when an existing one will do.

That these rules lead to a concise vocabulary is important — this makes it easy to combine commands to carry out several computer steps in a single utterance.

SLIDE 28 — Combining commands unleashes the power of speech

Window Close
Window Close No

3 Words
3 Words Bold

3 Lines
3 Lines Cut

This is a huge time-saver. Combining a pair of steps into one step is a 100 percent increase in efficiency. In general, if you don't have to think between steps, there's no reason to have separate steps.

One of the most important rules of human-machine grammar is that the command steps carried out by combined commands always follow the order of events.

This rule cuts out a lot of alternate wording possibilities and establishes a pattern, making it much easier to remember or guess how a command would be worded.

SLIDE 29

"Window Close No" hits the Close window command, then says no to the ensuing save dialog box **"3 Words Bold"** selects, then bolds the three words after the cursor, **"3 Lines Cut"** selects, then cuts the three lines below the cursor.

Another thing we pay attention to is visual and audio feedback. When you use the mouse to do an action that involves several separate steps, like selecting a paragraph, cutting the paragraph, moving the cursor to another location, then pasting the paragraph, you by default follow exactly what is happening.

We construct our speech commands to make sure that when you, for instance, select, cut, move and paste a paragraph using a single command, you're able to follow the action by seeing the paragraph highlighted in its original location before it is cut, then highlighted after it is pasted in the new location. It's important that this kind of feedback not become annoying, however, so it happens quickly.

As you watch the demonstration of Utter Command make sure to notice the types of words, their order, and the resulting patterns. Also be on the lookout for feedback.

Now I'll tell you a little more about Redstart Systems and Utter Command.

SLIDE 30

Redstart Systems is a company of people who use speech recognition — we think the key to making good software is using it.

Utter Command is our first product, and the foundation of a set of speech recognition interface products. We expect that Utter Command will be available this summer.

The Human-Machine Grammar will be posted on the Redstart Systems Web site in the form of a set of rules and dictionary of words by the time Utter Command becomes available. It contains all the rules and words used in Utter Command as well as all the words we use internally. The dictionary is an active document.

We are encouraging anyone who writes speech commands to use it, and we consider it one of our most important jobs to keep it updated.

Now we'll go through a short demo.

In this demo we'll prepare an email about a trip to Rome, and along the way you'll see a lot of the text handling and cut and paste type operations that are common in email and word processing tasks. Be sure to notice related words, word order, and patterns.

The first step of preparing an email is opening the email program and starting a message. For this step we'll show you how the NaturallySpeaking speech recognition engine works alone, then we'll show you what Utter Command speech interface software adds to the picture. The rest of the Demo will just show what Utter Command can do.

So, using NaturallySpeaking alone, this is how you would set up this email message. We are using Eudora, but other email programs like Outlook are similar.

Start Eudora
Message
New Message
Phil Comma Lisa
Tab Key
Rome trip
Tab Key
Karen
Tab Key
Tab Key
Phil Comma Lisa Comma
New Paragraph

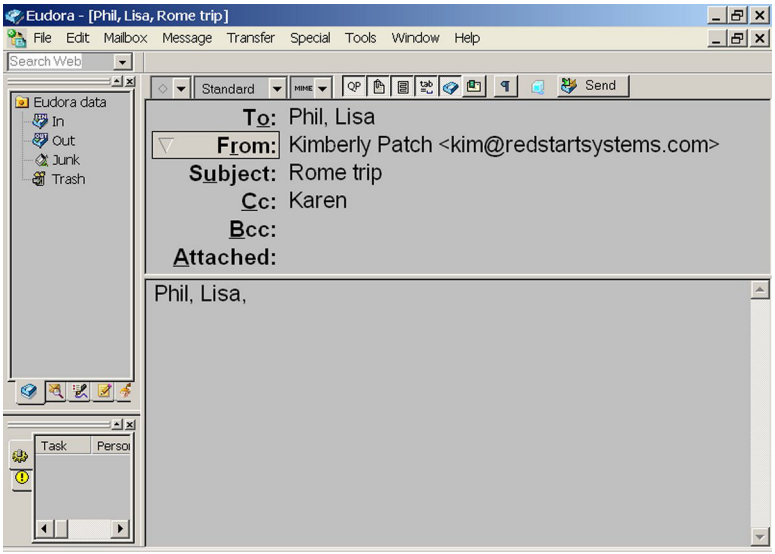
So that was 12 speech commands to get this far, and it took about 30 seconds.

Now we are going to close this window and do the same thing with Utter Command. Again, we are using Eudora. The commands are the same using Outlook.

To close the window we're going to use a combined command that allows you to close the window and say "no" to saving the changes in one step.

Window Close No

Eudora Phil Lisa CC Karen
Rome trip
Tab Times 3



That was 3 commands and about 10 seconds — three times as fast and four times fewer commands to wear your voice out on.

Let's get back to our trip:

Cap Hi guys Comma getting excited about the trip to Rome Question Mark

As you'll see during the next couple of minutes Utter Command allows you to move around in and change text fairly easily as you are writing.

2 Before
3 Befores Delete
1 Afters Bold
1 Left

Notice that this next command works even when the cursor is not at the end of the line.

Another Graph

Here's the brief itinerary Colon
Home Delete Cap Hotel

Another Graph

This next command is fairly advanced — it allows you to quickly paste common lists like days of the week or months of the year in text — these lists are also useful in spreadsheets

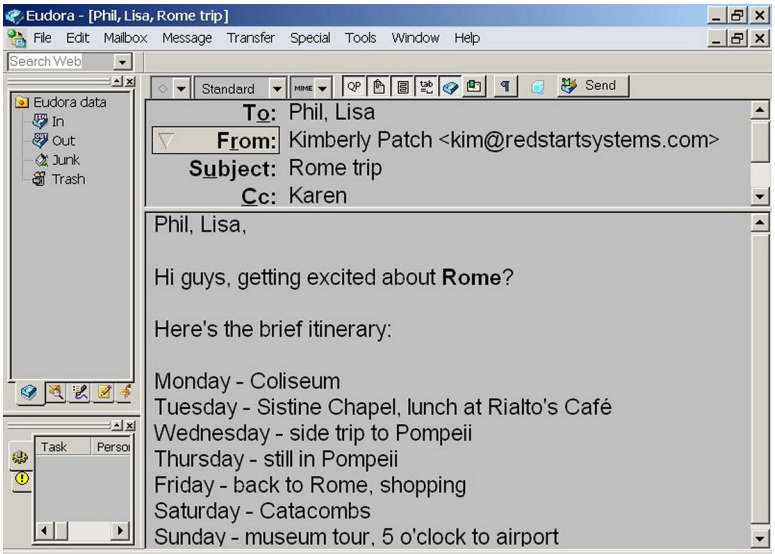
Days Enter Hyphen

In the next few commands Bill is going to fill in a schedule for a week in Rome. Notice that some of these commands combine two or more keystrokes. And keep in mind that every time you do two things using a single speech command it increases your efficiency by 100 percent.

7 Up End
Coliseum
Down End
Sistine Chapel, lunch at Rialto's cafe
Down End
side trip to Pompeii
Down End
still in Pompeii
Down End
back to Rome, shopping
Down End
Catacombs
Add Caps
Down End
museum tour, airport

Now you'll see some text-handling commands.

1 Before
5 o'clock to



Another Graph

Here's some background for Wednesday's tour
Dash a short history of Pompeii
Home Delete Cap Hotel
Fifth Word Italic
Another Graph

We're going to turn the rulers on now — this makes it easier to place the mouse using speech, which we'll see in a minute.

Rulers On

Pompeii site
I'm going to interrupt here to tell you more about the Sites List — Utter Command includes a list of Web sites that you can add to, and you can name the Sites on the list anything you want. We were able to call up the Pompeii site because it was on Bill's Sites list. The Sites List commands work with your default browser, and they work whether a browser is open or not.

Now we're going to peruse the Web site, then use the numbers on the rulers to move the arrow in order to select text.

Screen 2
Screen 3
Screen 4
Screen 2
64 by 15 (may change)
Drag 14 Down
This Copy

Bill's going to close the rulers using the icon on the taskbar.

Tray 1
Exit

This next command allows you to call up a program from the task bar.

Window 2

And Bill is going to paste the text from the Web site into his email message.

This Paste

And here are a couple of links you should look at Colon
Another Line

Then he will paste a couple of links from his sites list

Catacombs Path
Another Line
Pompeii Path
Another Graph

Cap See you at the airport
Dash 7 AM Dash don't be late Period
Home Delete Cap Sierra
Another Graph
Lisa Dash I'll pick you up at 6 AM Period
Another Graph

Hyphen Bill

This next command selects and moves a paragraph — notice that you see the paragraph highlighted before and after the move so you can follow the action.

2 Up
Graph Space 1 Back

Now we are going to attach a file using a key combination shortcut.

Control Hotel
Then we will paste the address of the file.

Rome Path
Enter

If Bill were to send this message he could use another key combination shortcut. Now we'll close the browser — the third window on the taskbar.

Window 3 Close
Window Close No

SLIDE 31

That is the end of this demo. I hope you enjoyed the talk and the demo, and I'm happy to answer any questions you might have. I should also mention that if you have comments, questions, criticisms or anything else to say about speech recognition and Human-Machine Grammar later, feel free to send email.