



at&t

MASH UP!

**A speech mashup framework for
mobile voice search**

**Jay Wilpon (jgw@research.att.com)
Giuseppe (Pino) Di Fabrizio**

AT&T Labs – Research





Challenge

- Existing business models in the speech industry limit scale of development community
 - Companies need to buy/integrated/support speech technology software, hardware, and network connectivity
- **Our Goals:**
 - Reduce barrier-of-entry (e.g., cost, ease of access to core speech technologies) for innovators
 - Grow innovators pool – create new service opportunities
 - Grow demand for data services

Web 2.0



Web 2.0 memes cloud as constructed by Markus Angermeier on 11/11/2005

Resources – Ubiquitous and Pervasive

- Network Connectivity
- Bandwidth
- Computing

Speech Mashup Vision

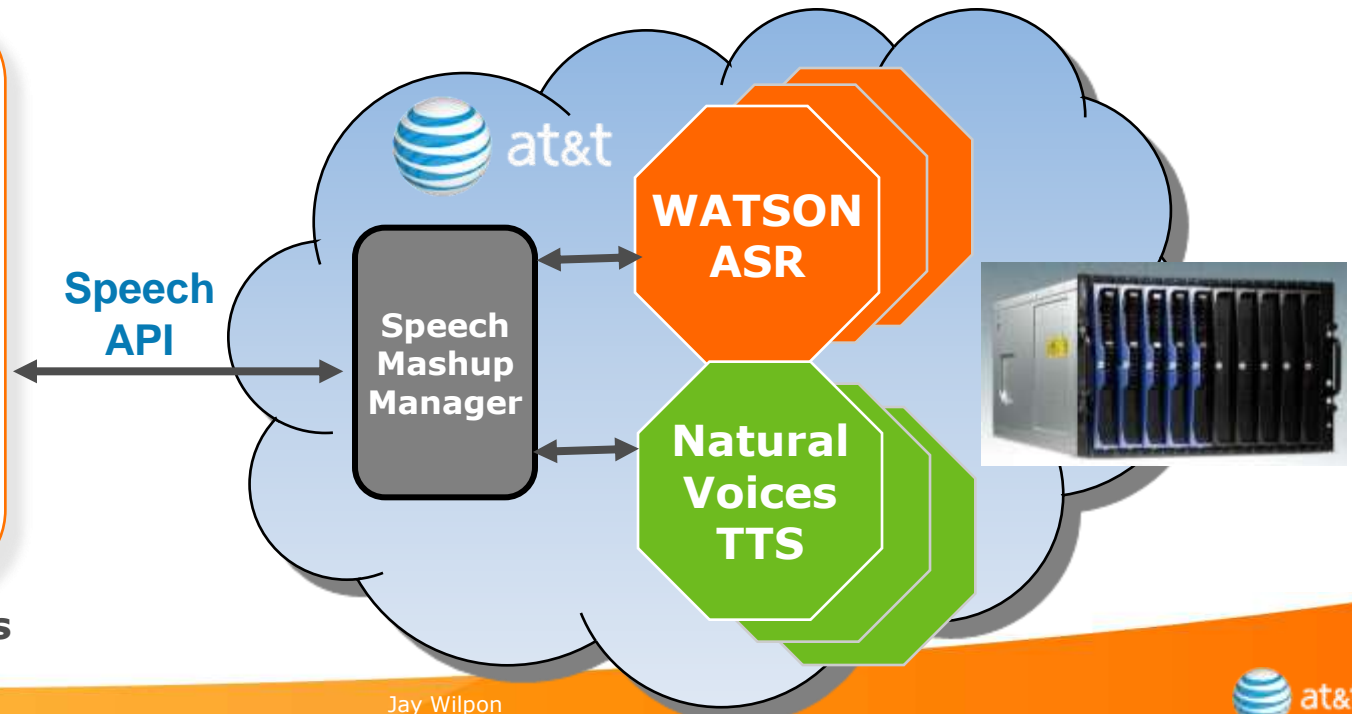
Grow development community; Enable Developers To Easily Create Voice-Enabled Services

Enable Best in Breed: Provide speech technologies and APIs, thus reducing high barrier-of-entry and enabling a new generation of service innovators

Benefit from Best in Breed Service Innovators: Many new services, many new vendors, increase customer base, new business models



Speech Mashup Clients





Speech Mashup – What is it

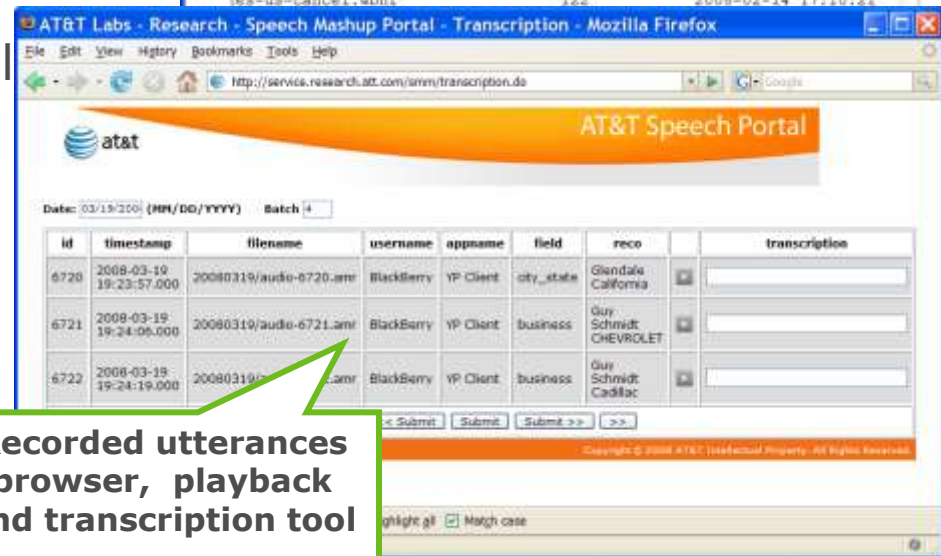
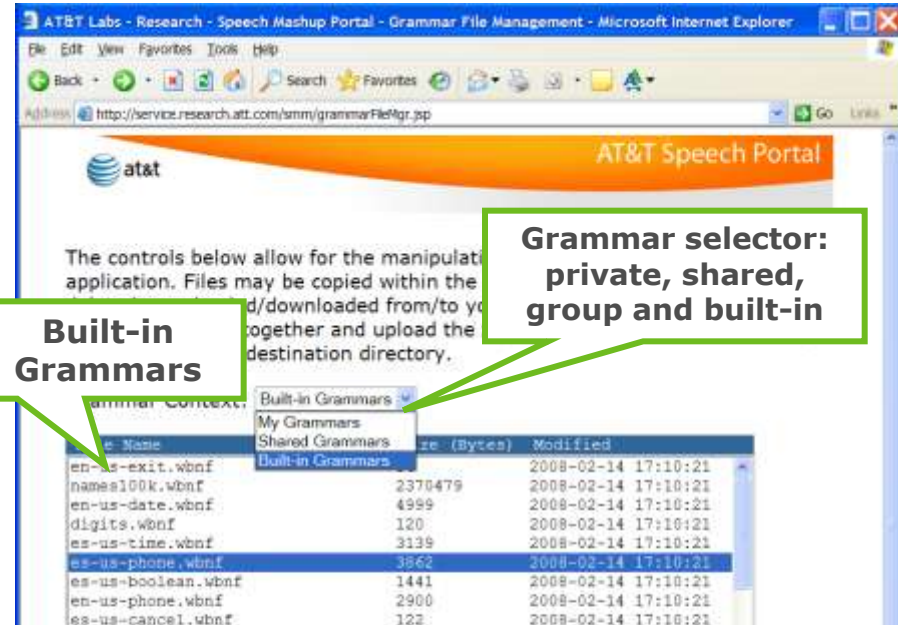
- Paradigm shift in technology/business model by integrating web services and speech technologies
- Publicly available web API for AT&T WATSON automatic speech recognition (speech to text) and AT&T Natural Voices (text to speech) synthesis
- “Voice-enabler” for mobile devices, including iPhones, Smartphones, or any network-enabled device with audio input (including PCs)
- Advanced multimodal interaction support by combining gestures and voice inputs

Application Enabling

- **Mobile**
 - Local business search (<http://www.speak4it.com>)
 - General internet voice search
 - Restaurant & hotel recommendations
 - Online ordering / transaction fulfillment
- **IPTV**
 - Movie on demand search
 - Media content search
 - Voice controlled DVR
 - Multimodal remote control
- **Multimedia Search/Indexing**
 - TV shows, movies, other video content
- **Security**
 - Voice Signature

AT&T Speech Mashup Portal

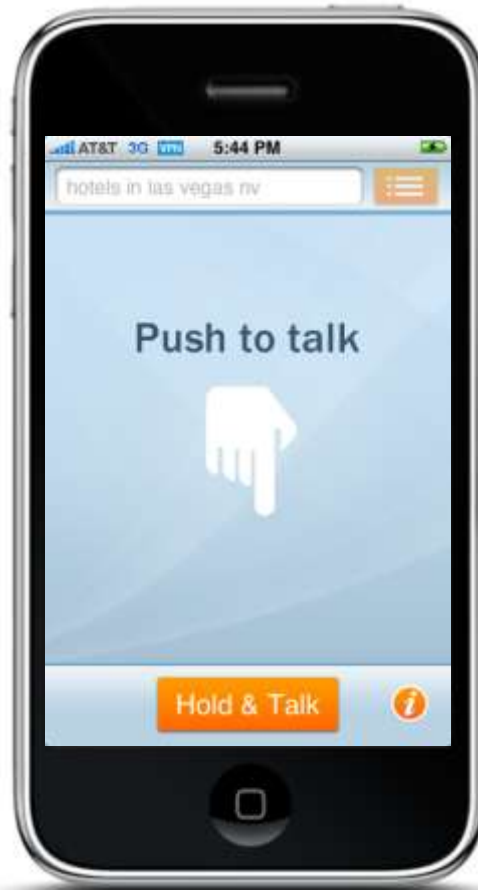
- User accounts management for internal and 3rd party application developers
- Direct access to speech mashups API for registered users
- Quota control (disk space and number of hits/day/account)
- Built-in set of basic grammars
- Web-based grammar compiler
- Grammars and language model sharing (public/private)
- Real-time data logger
- Utterance recorder and online transcription tools





Application Prototypes

Speak4it!



iPizza Speech Mashup Demo

Order pizza with the speech mashup

iPizza implements a multimodal interface for ordering pizza online. It combines speech input with graphical interaction on the iPhone to enable users to rapidly select menu items on a mobile device.

User's can speak naturally and request multiple items at the same time. The web interface allows easy navigation to update the items in the shopping cart.

1 Tap the Talk button.



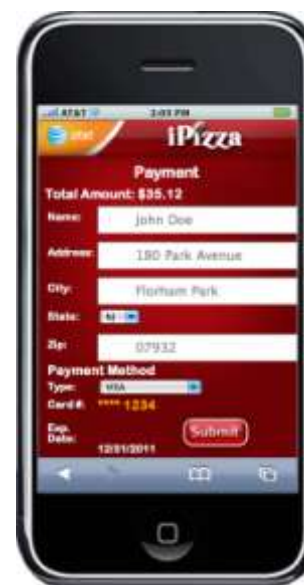
2 Say what you want using natural language: 'I'd like to order a pizza with cheese, mushrooms and ham, two diet Pepsi and backed cinnamon sticks'.



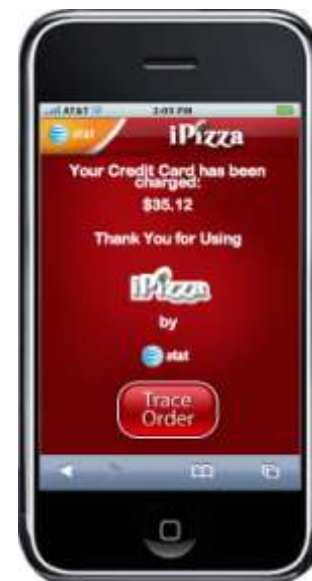
3 Complete your order.



4 Validate your order.



5 Trace it online.



iMOD – Multimodal Movies on Demand

