Problem

- Traditional IVRs have severe information limitations
- Limited self-service availability and usage
- Lengthy average handling times for limited progress
- Customers get lost in IVR menu tree
- Eventually customers transferred to a human agent, but still wait
- Many misrouted calls and Agent-to-Agent transfers
Problem

- Customer might be driving, or otherwise unable to look at a screen or type
- Live agents don't always have sufficient expertise in all necessary domains of interaction
- Subject to human emotion
- Leads to lower customer satisfaction
Solution
Solution: Branded Synthesized Agent

Call centers can dispatch a virtual agent

- Multi-style, multi-domain, multi-mood synthesized voice
- Call center can customize the voice to its preferences
- Including creating a branded voice
Multimodal Omnichannel

Interaction Channels

Phone | Mobile Apps | Web | SMS | IP/OTT Messaging Apps | Chatbot | Offline

Media

Voice | Voice, Text, GUI | Text, GUI | Text | Voice, Text, GUI | Voice, Text, GUI | Text

One Application, Multiple Channels, Optimized Interactions with Voice, Text, GUI & Rich Interactive Media
Branded Voice Solution

- Branded voice might be built from existing IVR recordings
- …or from new recordings of call center agents
- Can choose among various styles, domains, and preferences for normalization and speech gestures
- Fine-grained control over intonation, using mark-up language APIs, dictated prosody, and a number of other tools
- Can easily integrate with chatbots
Benefits to Customer

- can have longer interactions without sapping resources
- won't ever be asked to wait for the next agent
- should always be satisfied with the "expertise" of the agent in a number of different domains
- always have the option to wait for a human
- will never hear an "agent" angry or frustrated (unless desired)

In short: higher customer satisfaction
optimization of resources
Why Synthesized Voices
Why Synthesized Voices

• Allows for smooth, non-stitched audio (not pre-recorded, then TTS, then pre-recorded...)

• Also no need to record new prompts with voice talent. Just generate the prompts using the same TTS model

• Fine-grained control over pronunciation and intonation means the audio should sound exactly as you’d like it to
Why SpeechMorphing
Problems With Existing Synthesized Voice Offerings

- **Limited availability**: voices are limited to neutral and generic voices; typically limited to 2 per language; all vendors’ voices are indistinguishable and sound the same
- **Long time & high cost to build a specific voice**: typically requires months to build and over $50K to create a customized Voice
- **Low prosodic quality**: can be irritating and annoying when outside prosodic default, or when reading longer text
- **Limited features for 2nd tier language markets**: voices are limited to flat without being intentional and situational. Too much effort and too little data
Why SpeechMorphing

- Branded, expressive voices
- Require very little recording data, built quickly
- Natural conversational interactions:
  - Multiple moods and styles
  - Multiple domains
- Cutting-edge linguistic modeling, neural networks and advanced speech technologies Dynamic, scalable and economical Voice-on-Demand solution
- Complemented by customized voice building tools
- Can be integrated with any conversational application or service
Creating a Branded Voice
Creating a Branded Voice

- Half an hour of data, force-aligned and annotated with prosodic notation ToBI++ (using SpeechLabel tool, next slide)
- **Speaker adaptation**: The data is used as an extra layer in training a TTS voice, based on an existing larger corpus of data, which has been used to create other voices
- Less data is required for a new voice, because only certain aspects of the speaker’s voice quality need to be learned. Other aspects are speaker-independent, and already known
SpeechLabel: Prosodic Annotation
Creating a Branded Voice

• Need to map from existing model to speaker-dependent features of the target voice

• I.e., if I have sufficient examples of the sound [i] with pitch accent H* for target voice, then I can find a vector mapping from [i]+H* in my existing model to [i]+H* in the target voice

• When I generate TTS, I can use the existing models (duration, acoustic), plus those mapping vectors, to approximate my target voice
Front-End Management
Front-End Toolkit: Dictionary

<table>
<thead>
<tr>
<th>Dictionary</th>
<th>Compounds</th>
<th>Lexicon</th>
<th>Corpus</th>
<th>Prosody</th>
<th>Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>English</td>
<td>noun</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origin</td>
<td>English</td>
<td>not common</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Style</td>
<td>neutral</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acronym</td>
<td>not acronym</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Register</td>
<td>formal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammar Attributes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Gender: all
- Number: all
- Definiteness: both
- Declinability: regular

Proximity Rules | Translations
---|---

Semantic Relations
- Parameter
- Period
- Person
- Phenomenon
- Plant
- Potential
- Process
- Rank
- Relationship
- Relative
- Result
- Quantified

Semantic Groups
- Enh-Foot-1
- Enh-Foot-2
- Enh-Foot-3
- Enh-Foot-4
- Enh-Foot-5
- Enh-Foot-6
- Enh-Foot-7
- Enh-Foot-8
- Enh-Foot-9
- Enh-Foot-10
- Enh-Foot-11

Lexical Relations
- Army & warfare
- Astronomy
- Banking
- Biology
- Botany & Plants
- Business
- Chemistry & Materials
- Cinema
- Cognition
- Commerce
- Communication
- Computer
- Construction

Definition
- a linear unit equal to 12 inches or a third of a yard. "He is six feet tall!"

Etymology
- additive

© 2018 Speech Morphing, Inc.
Front-End Toolkit: Rules
Front-End Toolkit: Monitor

The image shows a screenshot of the Front-End Toolkit Monitor interface. It displays a list of options for different types of text normalization, such as POS tagging, frame annotation, and semantic group and fill annotation. The interface also includes a section for disambiguation, transcriptions, and a workspace for text editing. The window includes a list of words with their corresponding POS tags and semantic annotations.
Domain Management
Domain Manager
Domain Manager: Allowed Moods
Domain-Specific Pronunciation
## Domain-Specific Disambiguation and Normalization

### Image Description:

The image displays a software interface with a table and various text boxes. The table appears to contain domain-specific rules and attributes, possibly for language processing tasks.

**Table Columns:***
- **Domain**
- **Attributes**
- **Typed Rules**
- **Altered Types**
- **Altered Names**
- **Altered Signs**
- **Altered Units**
- **Word Replacement**
- **Word Normalization**
- **Disambiguation Rules**
- **Normalization Rules**
- **Precedence Rules**

**Sample Row:**
- **Domain:** Health
- **Attributes:** Age, Blood Pressure
- **Typed Rules:** 
  - *Type: Age*
  - *Name: Blood Pressure*
- **Precedence Rules:**
  - *Order: Age > Blood Pressure*

**Interface Elements:**
- **Username:** [Field]
- **Language:** [Dropdown]
- **Rule Set:** [Dropdown]
- **Selected Domain:** [Dropdown]

This interface is likely used for fine-tuning language processing rules specific to different domains, ensuring accurate disambiguation and normalization of terms and phrases.
Hi! I’m Anita’s avatar. Anita won’t be available for about 10 minutes, but I could help you now. Would you like my help, or would you rather wait for Anita? [cheerful]

I’d like your help.

Great! What can I do for you today? [cheerful]

I need to transfer some money to my checking account.
A Multi-modal Interaction

You requested a change of password. Is that correct? [business-like]

No, I asked to transfer money, from one account to another.

I’m sorry, I didn’t get that. What would you like to do? [apologetic]

I need to transfer money.
A Multi-modal Interaction

You requested a transfer. Is that correct? [business-like]

Actually, can I change my password?

I’m sorry. [apologetic]

Would you like to talk to the real Anita? [business-like]

No. I’d like your help.
Smorph: Speech Rendering Application

![Smorph: Speech Rendering Application](image-url)
Smorph: Prosodic Templates
Smorph: Prosodic Templates
SpeechMorphing’s Deep Learning-Based Solution

- **Easy Voice Building**: Fewer recordings, quick and low-cost build
- **Multiple Moods**: High conversational naturalness
- **Branding**: Original voice or adapted
- **Smart Prosody**: Dictated or predicted, supports lists and long paragraphs
- **Audio Effects and Gestures**: Using proprietary vocoder for natural situational enhancements
- **Delivery**: On Demand service, on premise, or packaged
- **Standard and Extended APIs**
Thank You!