The Allure and Power of Talking with the Machine

Presented by Emmett J. Coin
Industrial Poet
February 14, 2001
DCI-CRM Chicago

The Allure and Power of Talking with the Machine

Rationale —

Our tools have always been our friends.
And we like to talk to our friends.

We all talk to non-human things

- To things that we like
  - Pets
  - Plants

We often talk to machines

- To things that work for us
  - Cars
  - Televisions
- To things that work against us
  - Dice
  - Parking meters

Why do we do it?

- Do we even realize it?
- Are we embarrassed to admit it?
- How do we feel if we can't talk?
- We know they don't understand us.
- Maybe we just *need* to talk.
- Perhaps we do it for ourselves.
- It seems natural.

What will I talk about?

- A quick overview on the mechanics of speech
- A high level view of how speech technology works
- The challenges using speech technology for CRM/SFA
  - Examples from experiences

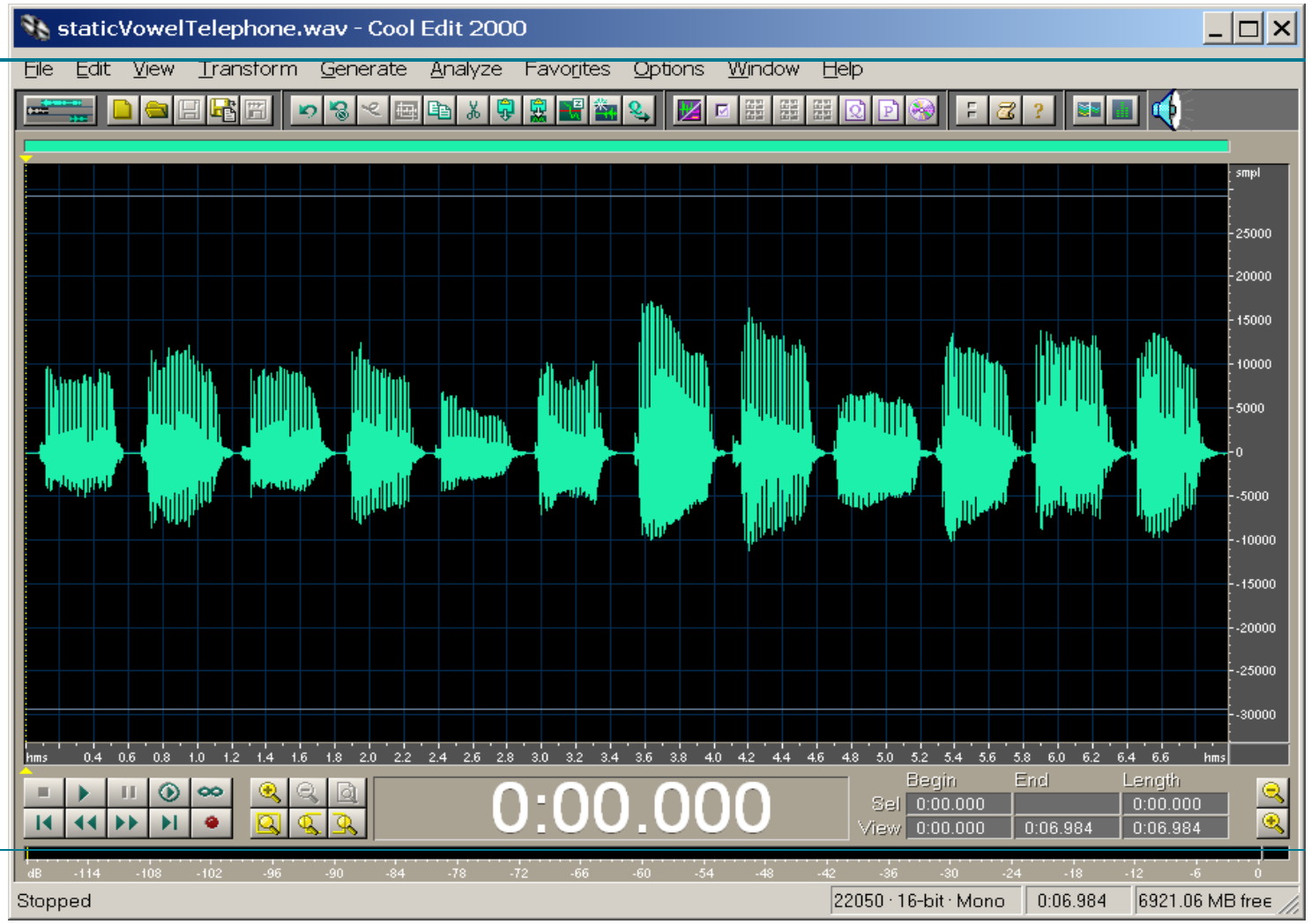
Speech is a defining characteristic of humans

- We detect speech patterns at a very early age (6 months)
- Speaking is *not* reading and writing
- Speech is inevitable, writing is not
- Speech may even be a fundamental need of humans

The Mechanics of Speech: A Tiny Primer

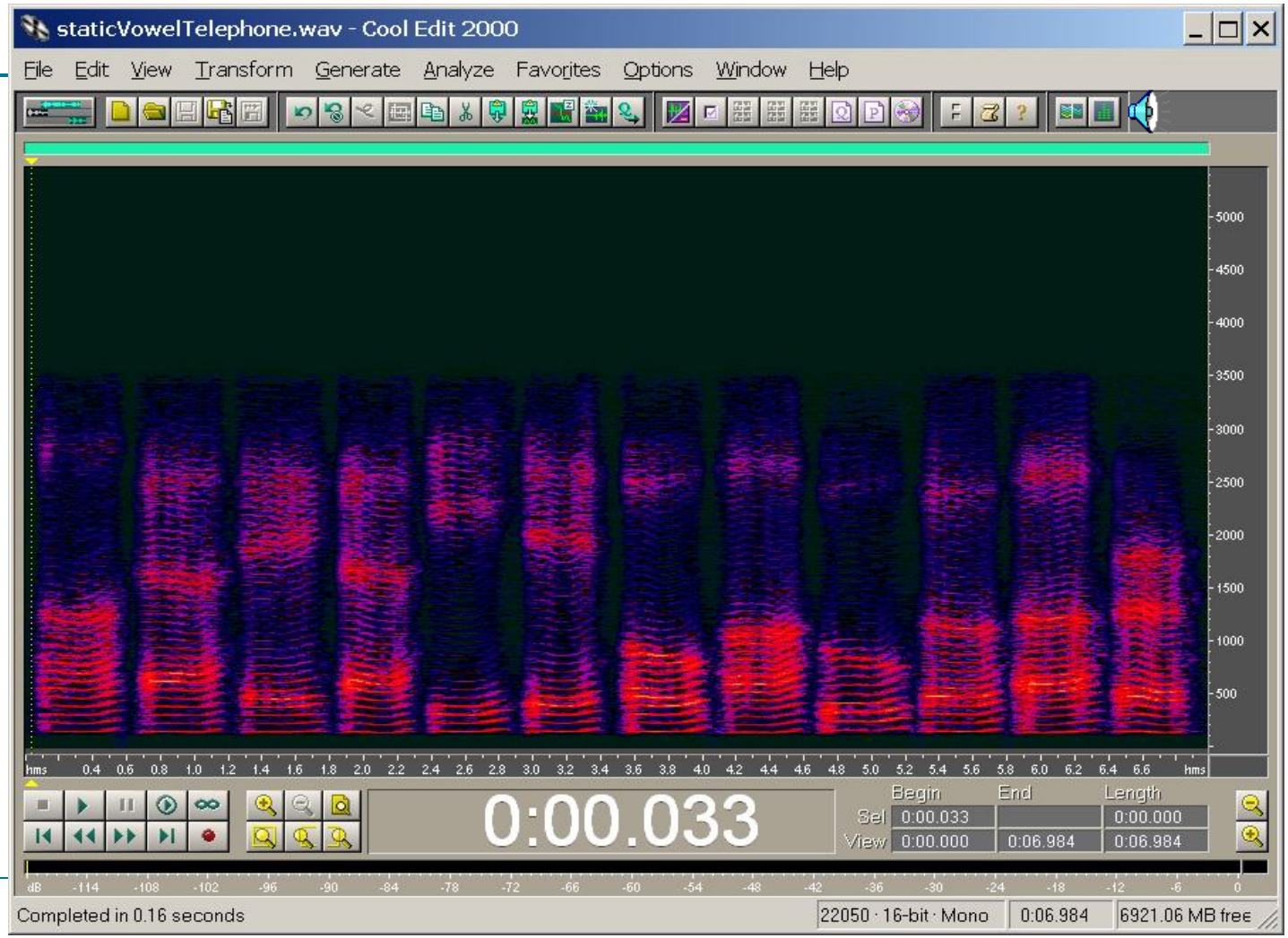
- Phonemes
  - Speech sounds
    - The Linguistics
- Formants
  - Power Spectrum
    - The Engineering
- Prosody
  - Pitch, Energy, Duration
    - The Music (and half the meaning)

Waveform:Regular Vowels



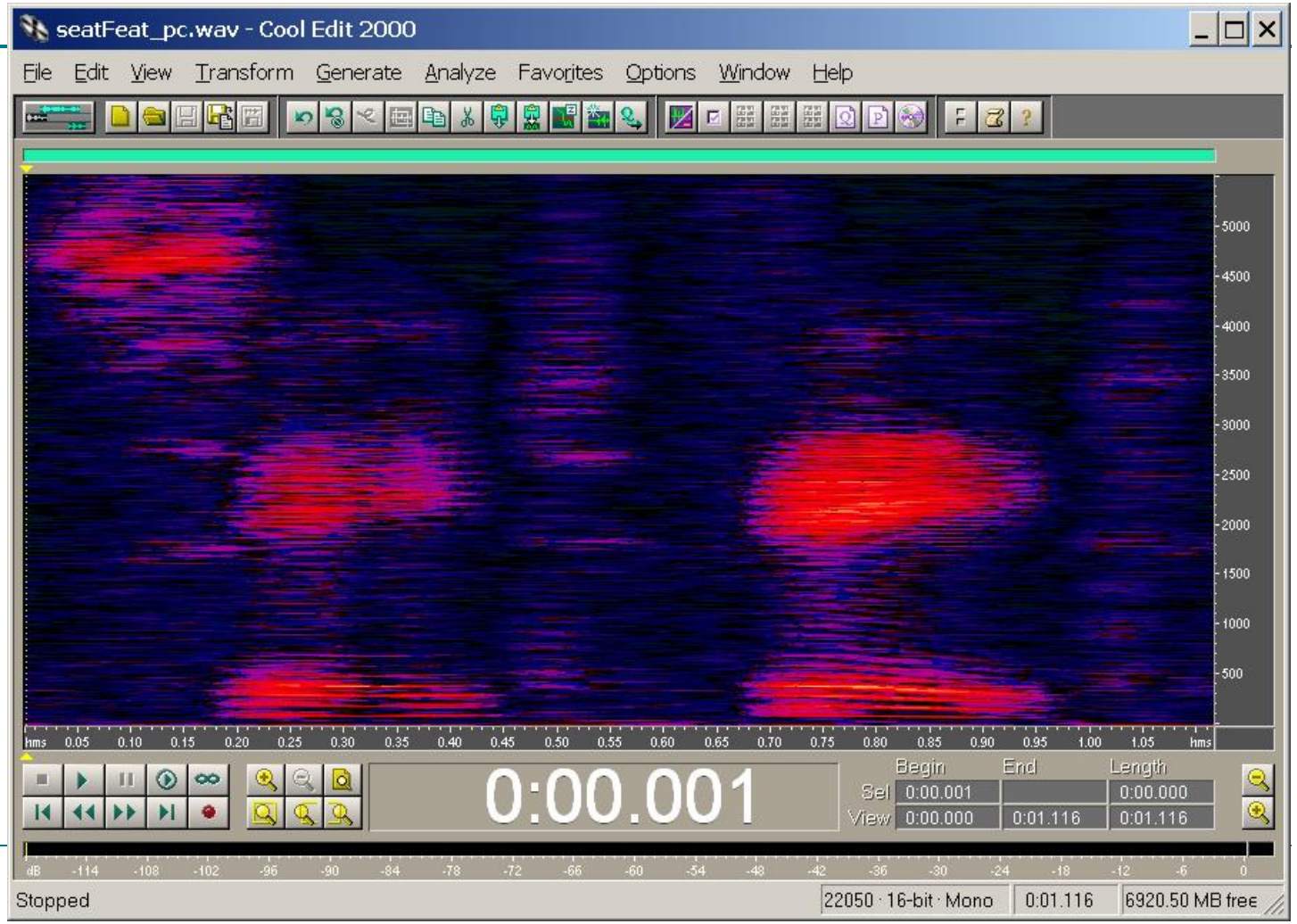
ej Talk

Power Spectrum:Regular Vowels



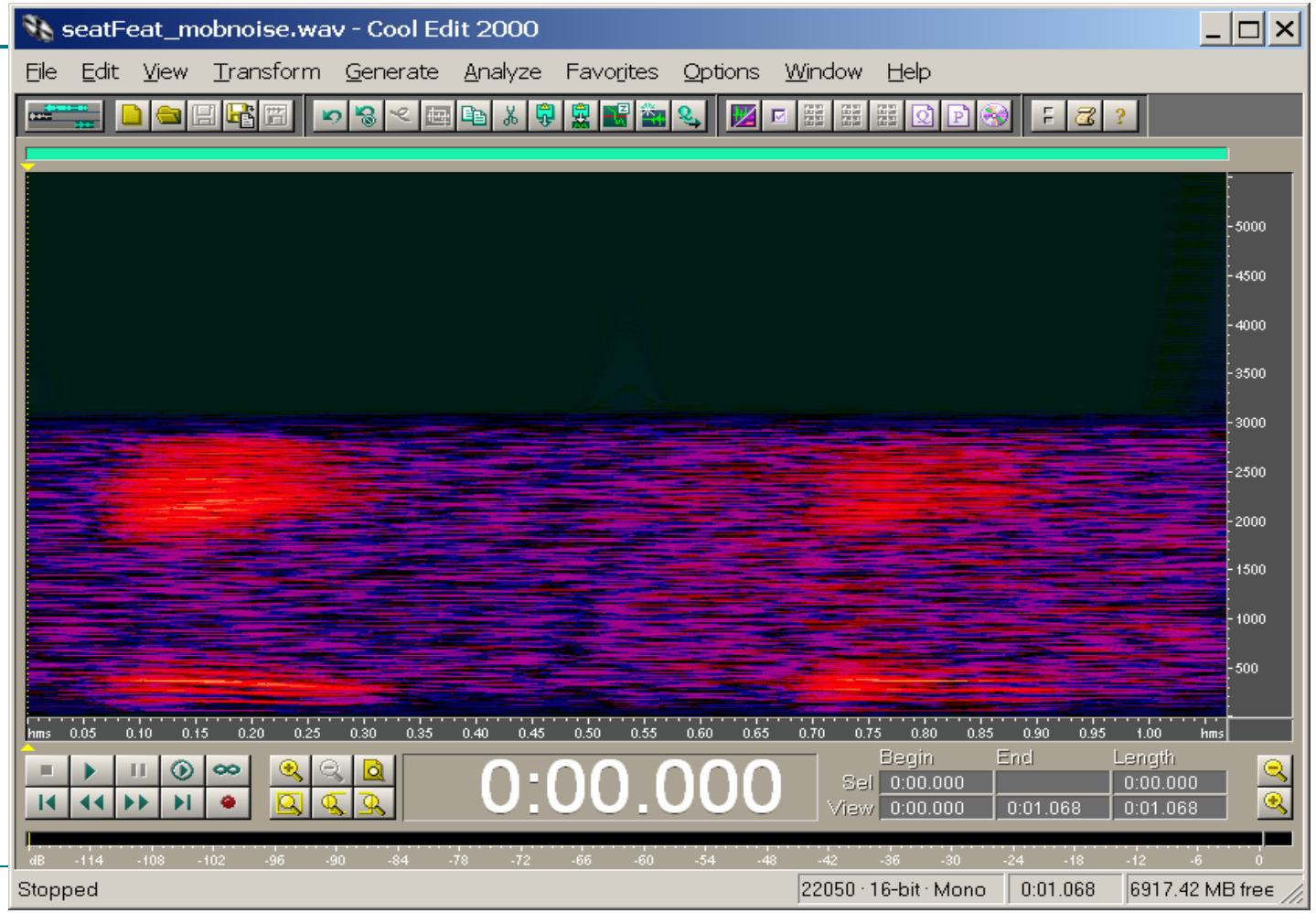
ej Talk

Radio Quality Speech



ej Talk

Typical Wireless Speech



ej Talk

What are the roots of human-computer speech technology?

- Mechanical artificial speech was demonstrated a century ago
- Linguists point out the “hard wired” nature of speech
- Computer scientists begin to think about detecting phonemes
- Computer performance gains in the 1960’s

Two problems, two solutions

- Listening
  - Detecting patterns
  - The machine ‘hears’
    - “CHICAGO ILLINOIS”
- Speaking
  - Making patterns
  - The machine ‘speaks’
    - “Please say the city and state.”

The Mechanics of ASR

(Automatic Speech Recognition)

- If we bend it, will it fit?
  - A warped sense of recognition
- HMM…
  - Hidden Markov Model
  - Beads on a string
    - Pick a bead
    - Make a pattern

Listening: Computer-Recognized Speech

- Template matching
- Phonetic
  - Finite Grammar
  - Statistical Grammar
- Training and speaker independence

Okay, we have speech
(Now we're talking, right?)

- What do all the words mean?
  - What did you think I meant?
  - What are you going to do about it?
- Different ways to attack the "understanding" problem.

Speaking: Computer-Generated speech

- Recorded schemes
  - Complete prompts
  - ‘Quilted’ speech
- Synthesized - Text To Speech (TTS)
  - From physical principles
  - Statistical mosaics of human speech

Synthesized Speech: TTS

- Formant-based
  - Physical principles
    - Vibration, resonance, etc.
- Concatenative (mosaics)
  - Sub phoneme
    - Transitions between phonemes
  - Large chunk
    - Snippets, words, phrases

Two “Tests” for CRM

- Recognition
  - What can the application hear?
- Response
  - How does the application speak?
- An example case: JustTalk

The Recognition Problem

- Finding existing information
  - Multiple choice problem
  - Relatively easy (you can guess)
- Capturing new information
  - Essay question
  - Relatively hard (guessing is bad)

The Response Problem

- Recorded human speech
  - Sounds the best (it *is* human)
  - Most limiting
    - Cannot handle novelty
- Synthesized speech
  - Easily identified as not human
  - Can say anything at anytime

The Choices that JustTalk made:

- Appeal vs. uniformity
- Different voices vs. one voice
- Implementation/server issues

What will make a success?

- Accommodate the user
  - Reduce the cognitive load
- Better than a human experience
  - Get more done
- Never fail
  - Your profit is in the automation

User Accommodation

- The simplest things aren't really
  - They are just the things we do with the least conscious thought.

More desirable than a human

- "*Please*, let me get their voice mail!"
- Buying an airline ticket online
- Using the ATM
- "I don't want to type this into my laptop later this evening!"

Never Fail

(At least don't appear to)

- Backup
  - Have humans save the day, but don't let on
- Collect
  - All relevant data on human backed incidents
- Incorporate
  - Continuously incorporate the lessons learned into the product

The JustTalk Approach

- Middle ground on accommodation
  - A wide range of phrasing instead of rigid command slots
- High ground
  - Aggressively reduce the “homework” component
  - “Ignore that man behind the curtain”
    - Do the right thing now and do it *automatically* later

Training for the user
(Does it help?)

- Yes
  - If users are trained enough
  - If they exercise the skills frequently
- No
  - If the skills are easily lost
  - If the cost of skill retention is more than the benefit

Summary

- Speech is our natural inclination
- The system must accommodate people
- Retrieving vs. Capturing
  - Big difference

Conclusion

- The industry is on the path to powerful and effective use of speech as an application interface
- Enhanced versions of this presentation is available on:
  - www.ejTalk.com
  - www.JustTalk.com
- Thank you.

ej
Talk