Integrating TTS and Recorded Speech

The best of both worlds

Marc Fabiani
The State of the Art

- You’ve come a long way, baby:
- Equivalency with recordings is possible
  - Recorded Speech    TTS
- Accuracy can exceed human ability with perfect consistency
  - Bexar, Texas
  - Bagehot
  - Natchitoches, Louisiana
- TTS Voices are available in many languages
- Easy to Use (text in -> speech out)
- Cost effective
So why isn’t TTS used everywhere?

- **Variety**
  - TTS offers at best a half dozen voices per language
  - Custom voices are expensive and time consuming
  - With recordings the choice is enormous

- **Expressivity**
  - TTS can recite but it can’t act

- TTS will address these issues eventually, but until it does recordings will continue to be used
What’s wrong with using recordings?

The Daily Record
What else is wrong with using recordings?

- Breaks the simplicity of Text in -> Speech out
- Complexifies application development
- Exposes platform and/or OS based inconsistencies
Why is this a big deal?

- **Software as a Service is the future of TTS**
- **Some characteristics of Software as a Service:**
  - Remotely Hosted, Network Accessed, Ubiquitous
  - Thin Client
  - Simple API
  - Transaction and/or subscription based
- **This is not a new idea**
Within the Reach of Millions

The most valuable things we touch are the intangible things gifts of modern man—sun, moon, millions of miles to the sky, green meadows and laughing islands, miles of human voices—tiny, lovely, kind, and friendly speech. Those of universal speech, some of the most useful are the telephone and the gramophone. There is no need here to mention them.

It is a great tribute to the value of the telephone that within a few short generations, it has come to be ranked among these current things, for truly are we a nation of millions of people. It unites and makes for easy and simplified living. It reduces the range of time very considerably. It allows one person, inaction, to send a message in less of a world. Daily it moves millions at once, rapidly, simply, saving time, money, and work. It multiplies earning power, new power for more income. Winds blow together the human fabric. Help the individual man and woman to triumph over the complacency of a vast world. Your voice without limits tells the world of your worth and universal thing as the telephone. You can only know that its value must be infinite.
Simple API
Transaction and/or Subscription Based

---

**Pennies For Wings**

The voice of a friend. Reassuring words from father, mother, son or daughter. A hurried call for aid in the night. You cannot put a price on such things as these.

Yet this is true—telephone service is cheap in this country. No other people get so much service, and such good and courteous service, at such low cost.

---

**Bell Telephone System**

The Bell System warmly invites you to visit its exhibit at the New York World’s Fair and the Golden Gate International Exposition, San Francisco.

-- JULY, 1933 --

Please mention Popular Science Monthly when writing to advertisers.
Solution

Replace this:

With this:
How?

- Custom Voices
- Partial Custom Voices
  - Provide pre-defined contexts for concatenating recordings
    - Dates
    - Times
    - Currency
    - Natural Number
    - Etc.
  - User simply drops in the recordings and the TTS engine does the rest
    - Text in -> Speech out
  - Allow for the integration of *any* recordings
Live Demo