

Mobile Voice

Conference 2015

*The Intelligent Connection*

The Sainte Claire Hotel San Jose, California - April 20-21, 2015

# Listening to Everyone

(or how to avoid a societal “speech divide”)

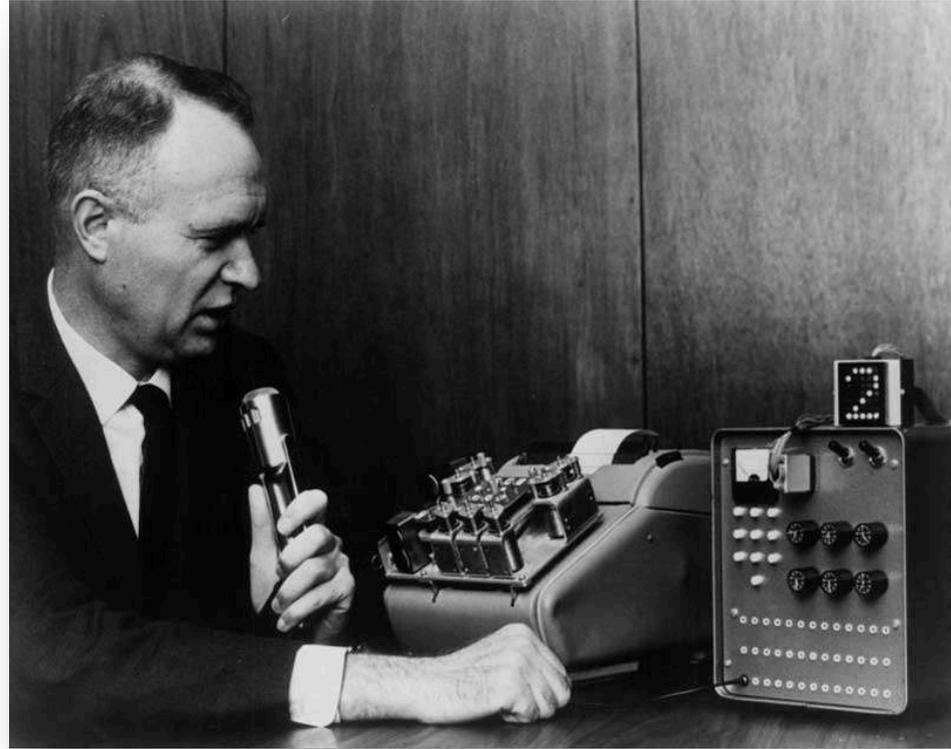
Marsal Gavaldà, PhD

@MarsalGavalda



Expect Labs

# The quest for speech recognition

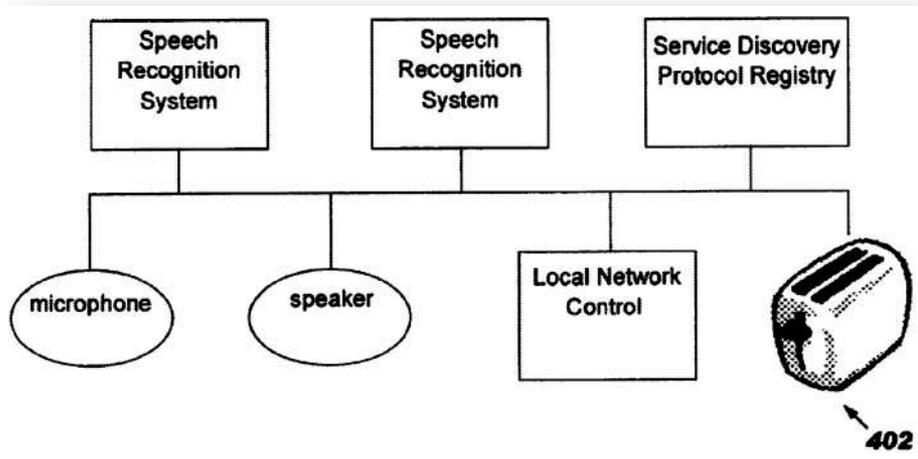


IBM Shoebox (1962)



"Her" (2013)

# The proverbial voice-enabled toaster



(12) **United States Patent**  
**Graham et al.**

(10) **Patent No.:** **US 6,654,720 B1**

(45) **Date of Patent:** **Nov. 25, 2003**

(54) **METHOD AND SYSTEM FOR VOICE CONTROL ENABLING DEVICE IN A SERVICE DISCOVERY NETWORK**

EP 1045374 10/2000 ..... G10L15/26

OTHER PUBLICATIONS

(75) Inventors: **Stephen Glen Graham**, Chapel Hill, NC (US); **Steven Michael Miller**, Apex, NC (US)

(73) Assignee: **International Business Machines Corporation**, Armonk, NY (US)

(\* ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

Rathi, S Blue Tooth Protocol Architecture (2000).  
<http://www.hometoys.com/htinews/dec98/reviews/voicetrek/voicetrek.htm>, "Voice Trek version 1.0", Dec. 1998 .  
<http://www.hometoys.com/htinews/aug97/reviews/homevoic/homevoic.htm>, "Home Voice Version 1.1", Aug. 1997 .  
 IEEE Transactions on Consumer Electronic, Aug. 1991, vol. 37, No. 3, pp. 395-400, "Solving automation problems using artificial intelligence", G. Evans.  
*Home Office Computing*, vol. 17, No. 2, p. 68, Feb. 1999, ISSN: 0899-7373, "Now You're Talking", D. Johnson .  
*Edge: Work-Group Computing Report*, vol. 8, p. 22, Mar. 3, 1997, "Voice processing: IBM's new voice-enabled Home Director puts IBM Aptiva to work around the house."

\* cited by examiner

(21) Appl. No.: **09/567,489**

(22) Filed: **May 9, 2000**

(51) Int. Cl.<sup>7</sup> ..... **G10L 21/00**

(52) U.S. Cl. .... **704/270; 704/270.1; 704/275**

(58) Field of Search ..... **704/270, 275, 704/270.1; 709/250, 310, 210**

*Primary Examiner*—Susan McFadden  
 (74) *Attorney, Agent, or Firm*—Gregory M. Doudnikoff, Carstens, Yee & Cahoon, LLP

(56) **References Cited**

**U.S. PATENT DOCUMENTS**

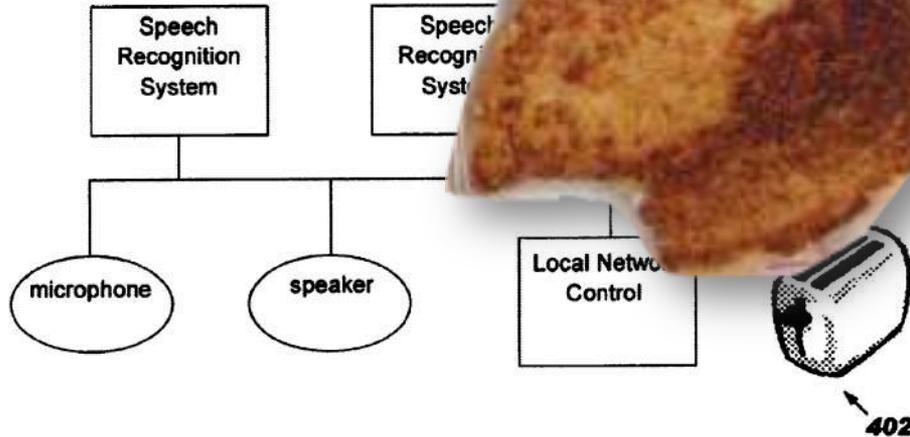
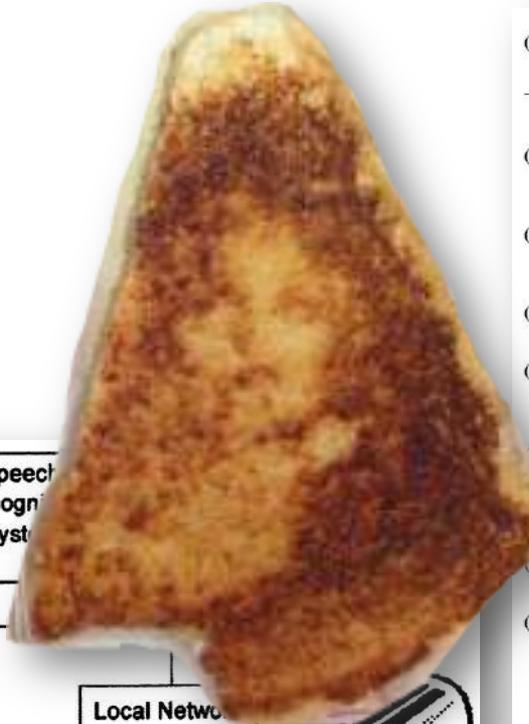
4,665,544 A	5/1987	Honda et al. ....	379/104
5,086,385 A	2/1992	Launey et al. ....	364/188
5,237,305 A	8/1993	Ishikuro et al. ....	340/286.01
5,579,221 A	11/1996	Mun ..... ..	364/188
5,621,662 A	4/1997	Humphries et al. ....	364/550
5,657,425 A	8/1997	Johnson ..... ..	395/2.84
5,878,394 A	3/1999	Muhling ..... ..	704/275
6,208,972 B1 *	3/2001	Grant et al. ....	704/275
6,405,261 B1 *	6/2002	Gaucher ..... ..	709/250
6,513,063 B1 *	1/2003	Julia et al. ....	709/219

(57) **ABSTRACT**

Devices are registered within a zone using a service discovery protocol. At the time of registration, a device sends a grammar and a set of help sequences to a control unit in the zone to be merged into a composite grammar and composite set of help sequences. If the voice command issued by a user is recognized, an appropriate method invocation is sent to the device to perform the desired operation. If the device in the voice command is recognized but the operation is not recognized, then the user is provided with a help sequence on how to properly perform operations with the device. If the device is not recognized, the command is forwarded to another zone within a set of zones until either the device is found or no such device is located.

**FOREIGN PATENT DOCUMENTS**

# The proverbial voice-enabled toaster



(12) **United States Patent**  
**Graham et al.**

(10) Patent No.: **US 6,654,720 B1**  
(45) Date of Patent: **Nov. 25, 2003**

(54) **METHOD AND SYSTEM FOR VOICE CONTROL ENABLING DEVICE IN A SERVICE DISCOVERY NETWORK**

EP 1045374 10/2000 ..... G10L/15/26

OTHER PUBLICATIONS

(75) Inventors: **Stephen Glen Graham**, Chapel Hill, NC (US); **Steven Michael Miller**, Apex, NC (US)

Rathi, S Blue Tooth Protocol Architecture (2000).  
http://www.hometoys.com/htnews/dec98/reviews/voice-trek/voicetrek.htm, "Voice Trek version 1.0", Dec. 1998 .  
http://www.hometoys.com/htnews/aug97/reviews/home-voic/homevoic.htm, "Home Voice Version 1.1", Aug. 1997 .  
IEEE Transactions on Consumer Electronic, Aug. 1991, vol. 37, No. 3, pp. 395-400, "Solving automation problems using artificial intelligence", G. Evans.  
*Home Office Computing*, vol. 17, No. 2, p. 68, Feb. 1999, ISSN: 0899-7373, "Now You're Talking", D. Johnson .  
*Edge: Work-Group Computing Report*, vol. 8, p. 22, Mar. 3, 1997, "Voice processing: IBM's new voice-enabled Home Director puts IBM Aptiva to work around the house. "

(73) Assignee: **International Business Machines Corporation**, Armonk, NY (US)

(\* ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

\* cited by examiner

(21) Appl. No.: **09/567,489**

(2) Filed: **May 9, 2000**

(3) Int. Cl.<sup>7</sup> ..... **G10L 21/00**

(2) U.S. Cl. .... **704/270; 704/270.1; 704/275**

(38) Field of Search ..... **704/270, 275, 704/270.1; 709/250, 310, 210**

*Primary Examiner*—Susan McFadden  
(74) *Attorney, Agent, or Firm*—Gregory M. Doudnikoff, Carstens, Yee & Cahoon, LLP

(56) **References Cited**

**U.S. PATENT DOCUMENTS**

4,665,544 A	5/1987	Honda et al. ....	379/104
5,086,385 A	2/1992	Launey et al. ....	364/188
5,237,305 A	8/1993	Ishikuro et al. ....	340/286.01
5,579,221 A	11/1996	Mun ..... ..	364/188
5,621,662 A	4/1997	Humphries et al. ....	364/550
5,657,425 A	8/1997	Johnson ..... ..	395/2.84
5,878,394 A	3/1999	Muhling ..... ..	704/275
6,208,972 B1 *	3/2001	Grant et al. ....	704/275
6,405,261 B1 *	6/2002	Gaucher ..... ..	709/250
6,513,063 B1 *	1/2003	Julia et al. ....	709/219

(57) **ABSTRACT**

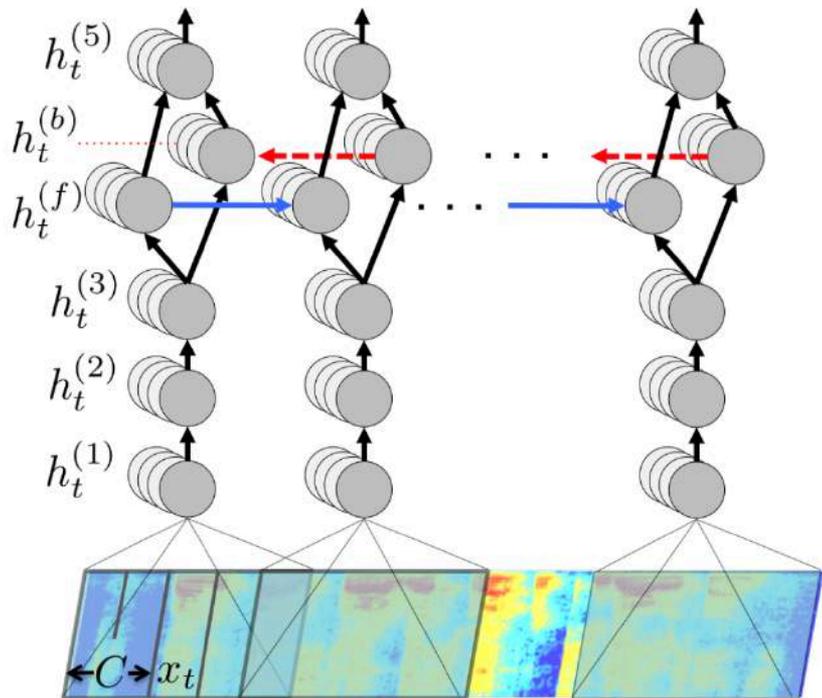
Devices are registered within a zone using a service discovery protocol. At the time of registration, a device sends a grammar and a set of help sequences to a control unit in the zone to be merged into a composite grammar and composite set of help sequences. If the voice command issued by a user is recognized, an appropriate method invocation is sent to the device to perform the desired operation. If the device in the voice command is recognized but the operation is not recognized, then the user is provided with a help sequence on how to properly perform operations with the device. If the device is not recognized, the command is forwarded to another zone within a set of zones until either the device is found or no such device is located.

**FOREIGN PATENT DOCUMENTS**

# Rapid adoption of speech recognition



# Prodigious improvements in accuracy

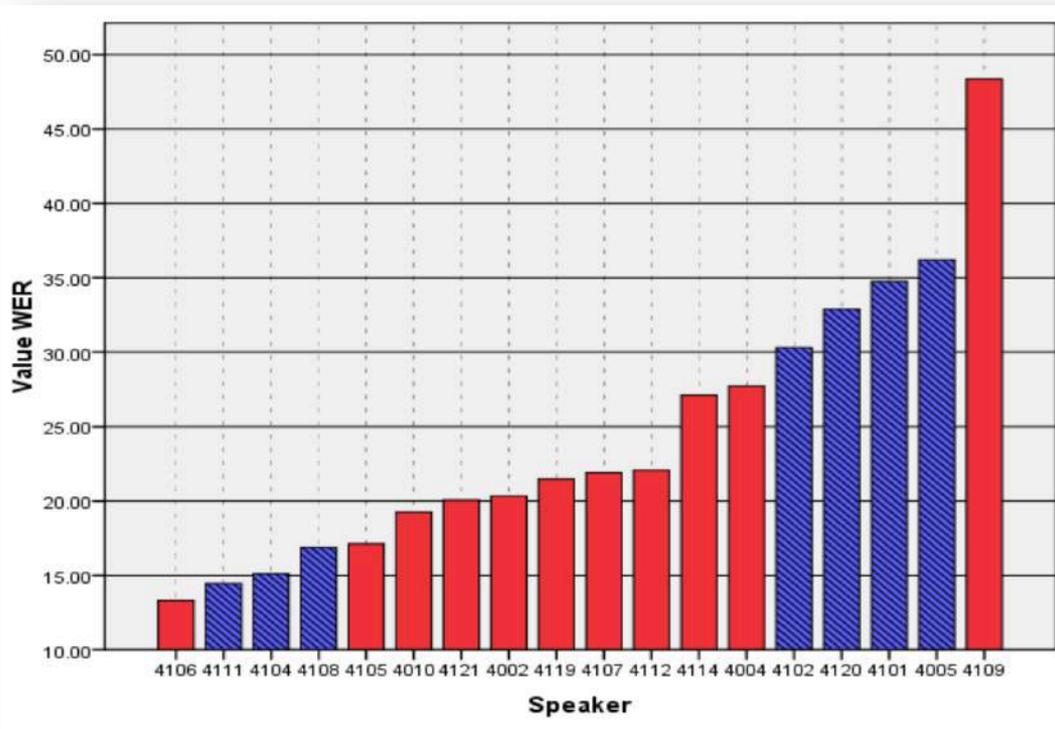


Model	SWB	CH	Full
Vesely et al. (GMM-HMM BMMI) [44]	18.6	33.0	25.8
Vesely et al. (DNN-HMM sMBR) [44]	12.6	24.1	18.4
Maas et al. (DNN-HMM SWB) [28]	14.6	26.3	20.5
Maas et al. (DNN-HMM FSH) [28]	16.0	23.7	19.9
Seide et al. (CD-DNN) [39]	16.1	n/a	n/a
Kingsbury et al. (DNN-HMM sMBR HF) [22]	13.3	n/a	n/a
Sainath et al. (CNN-HMM) [36]	11.5	n/a	n/a
Soltau et al. (MLP/CNN+I-Vector) [40]	<b>10.4</b>	n/a	n/a
<b>Deep Speech SWB</b>	20.0	31.8	25.9
<b>Deep Speech SWB + FSH</b>	12.6	<b>19.3</b>	<b>16.0</b>

System	Clean (94)	Noisy (82)	Combined (176)
Apple Dictation	14.24	43.76	26.73
Bing Speech	11.73	36.12	22.05
Google API	6.64	30.47	16.72
wit.ai	7.94	35.06	19.41
<b>Deep Speech</b>	<b>6.56</b>	<b>19.06</b>	<b>11.85</b>

Awni Hannun *et al.* "DeepSpeech: Scaling up end-to-end speech recognition." arXiv:1412.5567 (2014).

# But not for everyone?

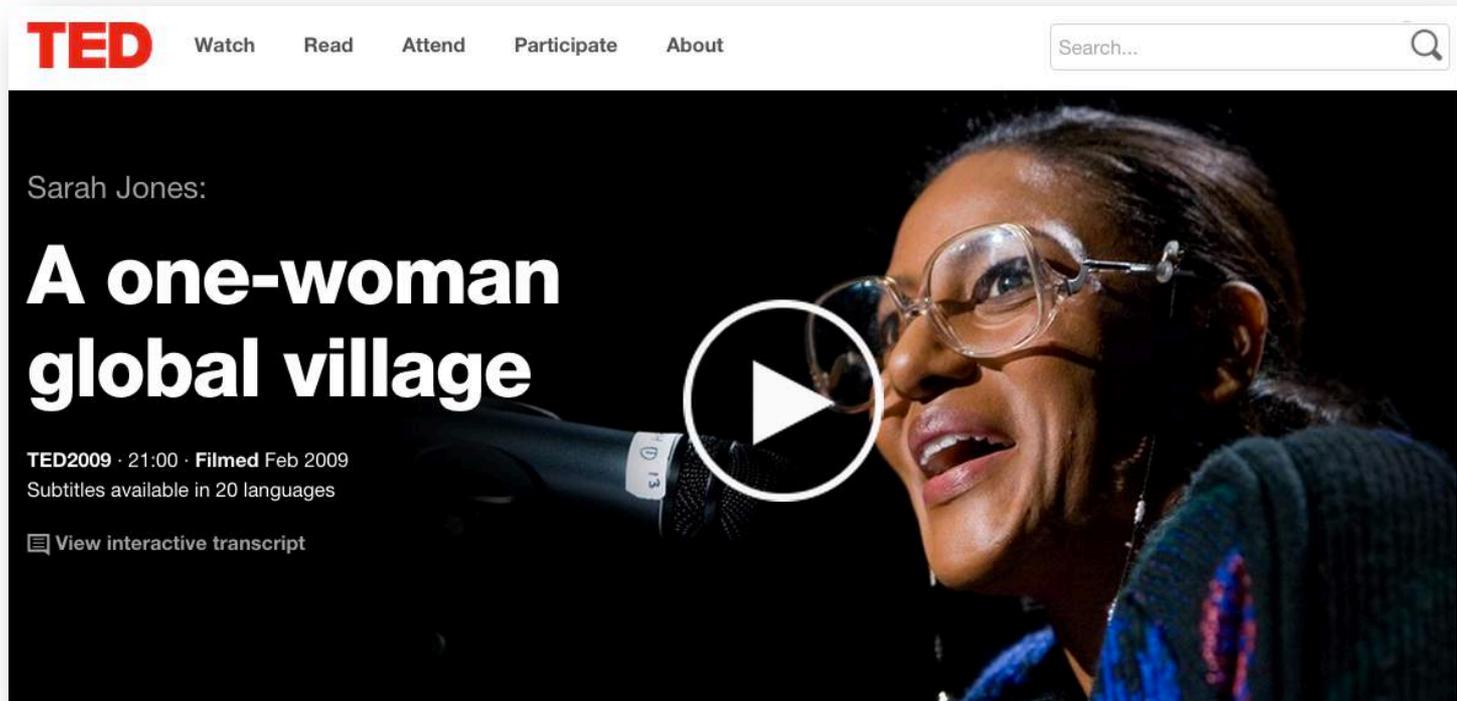


Beňuš *et al.* "Semi-automatic approach to ASR errors categorization in multi-speaker corpora." Slovko 2011.

Gender	WER(%)
all	35.3
M	33.5
F	36.0
Age range	WER(%)
60-65	29.1
66-70	28.1
71-75	36.1
76-80	45.1
81-85	41.0
86-90	54.9

Pellegrini *et al.* "Impact of age in ASR for the elderly: preliminary experiments in European Portuguese." IberSPEECH 2012.

# Accuracy: **Accented Speech**



The image is a screenshot of a TED talk page. At the top left is the TED logo in red. To its right are navigation links: Watch, Read, Attend, Participate, and About. On the far right is a search bar with the text 'Search...' and a magnifying glass icon. The main content area features a large black background with a close-up photograph of Sarah Jones speaking into a microphone. She is wearing glasses and has a joyful expression. A large white play button icon is overlaid on the center of the image. To the left of the image, the text reads: 'Sarah Jones:', followed by the title 'A one-woman global village' in large white font. Below the title, it says 'TED2009 · 21:00 · Filmed Feb 2009' and 'Subtitles available in 20 languages'. At the bottom left, there is a link that says 'View interactive transcript' with a document icon.

**TED** Watch Read Attend Participate About

Search...

Sarah Jones:

## A one-woman global village

TED2009 · 21:00 · Filmed Feb 2009  
Subtitles available in 20 languages

[View interactive transcript](#)

# Natural



to what extent do we self construct **dewey** self invent and  
how do we self-identify and how mutable is that identity  
**but** what if one could be anyone at any time

# Jewish



but anyway I should have started by **Shane** just how lucky  
I feel **a chechen** I opening experience **should** be **hiya**

# Indian



**on that** I must attend these meetings mingling with some  
**Thai some** politicians trying to make me feel comfortable by  
telling how often they are **leading** the **Cutty** in Georgetown

# Chinese



**Honda** we tried to teach our kids **r2p prolyy up that** tradition  
**about this** very **hot** you know as immigrant

# Jordanian

A woman wearing a white headscarf is shown in profile, speaking into a black microphone. The background is dark. The text is overlaid on the bottom left of the image.

I like to believe **but** for every **Wed** intended to rendered **ass**  
**death** to one another there **its** always a **leaky** connecting  
**he is** and **hides** across the continent's in **there**

# Accuracy: **Accented speech**

<b>Character</b>	<b>% WER</b>	<b>x worse</b>
Natural	21.3	1.0
Indian	29.9	1.4
Jewish	30.5	1.4
Chinese	32.4	1.5
French	35.1	1.7
Dominican	39.2	1.8
Jordanian	46.9	2.2

# Accented English: **World population**

## **Native:** 400M

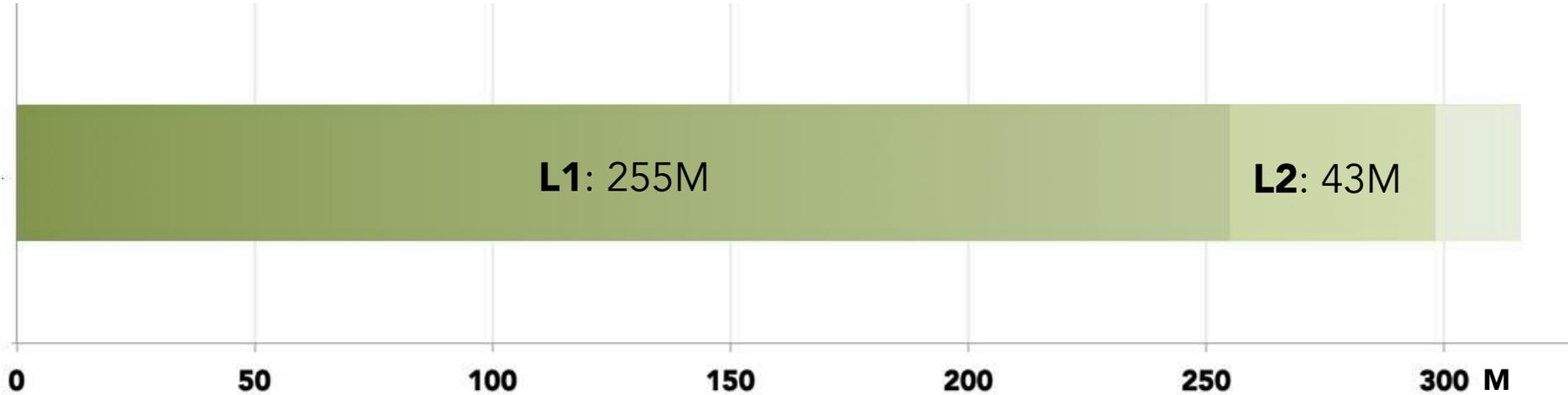
- Australia
- Canada
- UK
- USA
- New Zealand

## **Official:** 400M

- Bangladesh
- Ghana
- India
- Kenya
- Malaysia
- Nigeria
- Pakistan
- Philippines
- Singapore
- Sri Lanka
- Tanzania
- Zambia

## **ESL speakers elsewhere:** 400M

# Accented English: **US population**



# Analogy: Clinical trials

One Hundred Third Congress  
of the  
United States of America

AT THE FIRST SESSION

*Begun and held at the City of Washington on Tuesday,  
the fifth day of January, one thousand nine hundred and ninety-three*

An Act

To amend the Public Health Service Act to revise and extend the programs of the National Institutes of Health, and for other purposes.

*Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,*

**SECTION 1. SHORT TITLE; TABLE OF CONTENTS.**

(a) **SHORT TITLE.**—This Act may be cited as the “National Institutes of Health Revitalization Act of 1993”.

(b) **TABLE OF CONTENTS.**—The table of contents for this Act is as follows:

Sec. 1. Short title; table of contents.

**TITLE I—GENERAL PROVISIONS REGARDING TITLE IV OF PUBLIC HEALTH SERVICE ACT**

**Subtitle A—Research Freedom**

**PART I—REVIEW OF PROPOSALS FOR BIOMEDICAL AND BEHAVIORAL RESEARCH**

Sec. 101. Establishment of certain provisions regarding research conducted or supported by National Institutes of Health.

**PART II—RESEARCH ON TRANSPLANTATION OF FETAL TISSUE**

Sec. 111. Establishment of authorities.

Sec. 112. Purchase of human fetal tissue; solicitation or acceptance of tissue as directed donation for use in transplantation.

Sec. 113. Nullification of moratorium.

Sec. 114. Report by General Accounting Office on adequacy of requirements.

**PART III—MISCELLANEOUS REPEALS**

Sec. 121. Repeals.

**Subtitle B—Clinical Research Equity Regarding Women and Minorities**

**PART I—WOMEN AND MINORITIES AS SUBJECTS IN CLINICAL RESEARCH**

Sec. 131. Requirement of inclusion in research.

Sec. 132. Peer review.

Sec. 133. Inapplicability to current projects.

**PART II—OFFICE OF RESEARCH ON WOMEN'S HEALTH**

Sec. 141. Establishment.

**PART III—OFFICE OF RESEARCH ON MINORITY HEALTH**

Sec. 151. Establishment.

## Subtitle B—Clinical Research Equity Regarding Women and Minorities

### PART I—WOMEN AND MINORITIES AS SUBJECTS IN CLINICAL RESEARCH

#### SEC. 131. REQUIREMENT OF INCLUSION IN RESEARCH.

Part G of title IV of the Public Health Service Act, as amended by section 101 of this Act, is amended by inserting after section 492A the following section:

“INCLUSION OF WOMEN AND MINORITIES IN CLINICAL RESEARCH

“SEC. 492B. (a) REQUIREMENT OF INCLUSION.—

“(1) IN GENERAL.—In conducting or supporting clinical research for purposes of this title, the Director of NIH shall, subject to subsection (b), ensure that—

“(A) women are included as subjects in each project of such research; and

“(B) members of minority groups are included as subjects in such research.

# Analogy: Clinical trials

**I'M IN?** CLINICAL TRIALS for Better Health

LOGIN

New to I'm In? [Sign-up here.](#)

WHAT'S I'M IN? DIVERSITY READY TO JOIN? ENGAGEMENT NETWORK NEWS BLOG

Are you in?

Without the patients who volunteer to participate in clinical trials, the development of new treatments would not be possible.

LEARN MORE

f

t

I'm a Patient \_\_\_\_\_.

PATIENT

ADVOCATE

PHYSICIAN

CLINICAL TRIAL SPONSOR

CLINICAL TRIAL INVESTIGATOR

FRIENDS AND FAMILY

Each day, new treatments are being developed to address the health challenges facing diverse communities, including African American, Asian American and Hispanic populations.

Unfortunately, these communities are underrepresented in the clinical trials that test the effectiveness of treatments. We can change that. [Are you in?](#)

**EMPACT** Enhancing Minority Participation in Clinical Trials

Home About Training Courses Our Progress News

Resources Directory Contact Us

Search

**Our Goal**

The mission of EMPACT is to increase recruitment and retention of racial/ethnic minorities into therapeutic clinical trials with the ultimate goal of reducing cancer-related health disparities.

**EMPACT's Goal**

[Click to watch a video presentation about EMPACT goals and vision.](#)

FEATURED TRAINING COURSE

**Successful Models for Minority Recruitment**

Based on Dr. Jean Ford's model.

FEATURED COURSE MATERIAL

**Defining and Identifying Cultural Factors Affecting Recruitment**

FEATURED NEWS

**Enrollment in Clinical Trials Improves Survival**

OUR FUNDER

NIH National Institute on Minority Health and Health Disparities

THE CONSORTIUM

UNIVERSITY OF MINNESOTA

UAB THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

THE UNIVERSITY OF TEXAS MD Anderson Cancer Center

UC DAVIS COMPREHENSIVE CANCER CENTER

JOHNS HOPKINS UNIVERSITY

# Call to Action: **Narrow the Speech Divide!**

## **Problem**

- The accuracy of speech recognition varies widely from person to person.
- It tends to perform poorly for children, seniors, and non-native speakers.
- We can't let large swaths of the population be cut off from the growing benefits of speech-based services.
- In fact, we should worry about disenfranchised speakers filing disability discrimination and civil rights class-action lawsuits.

## **Solution**

- Start with a more transparent characterization of the problem.
- Make concerted effort to improve accuracy for children, seniors, and accented speakers.
  - Include more "minorities" in the training datasets.
  - Make it a research and product priority.
- Wonderful chance to become better than humans!