

Lessons for Responsible Science from DARPA's Programs in Human Language Technology

Mark Liberman
University of Pennsylvania

<http://ling.upenn.edu/~myl>

The story begins in the 1960s...

... with two bad reviews by John Pierce,
an executive at Bell Labs
who invented the word “transistor”
and supervised development
of the first communications satellite.



In 1966, John Pierce chaired the
“Automatic Language Processing Advisory Committee” (ALPAC)
which produced a report to the National Academy of Sciences,
Language and Machines: Computers in Translation and Linguistics

And in 1969,
he wrote a letter to the Journal of the Acoustical Society of America,
“Whither Speech Recognition”

The ALPAC Report

MT in 1966 was not very good, and ALPAC said diplomatically that

“The Committee cannot judge what the total annual expenditure for research and development toward improving translation should be. However, it should be spent hardheadedly toward important, realistic, and relatively short-range goals.”

In fact, U.S. MT funding went essentially to zero for more than 20 years.

The committee felt that science should precede engineering in such cases:

“We see that the computer has opened up to linguists a host of challenges, partial insights, and potentialities. We believe these can be aptly compared with the challenges, problems, and insights of particle physics. Certainly, language is second to no phenomenon in importance. And the tools of computational linguistics are considerably less costly than the multibillion-volt accelerators of particle physics. The new linguistics presents an attractive as well as an extremely important challenge.”

John Pierce's views
about automatic speech recognition
were similar to his opinions about MT.

And his 1969 letter to JASA,
expressing his personal opinion,
was much less diplomatic
than that 1966 N.A.S. committee report....

“Whither Speech Recognition?”

“... a general phonetic typewriter is simply impossible unless the typewriter has an intelligence and a knowledge of language comparable to those of a native speaker of English.”

“Most recognizers behave, not like scientists, but like mad inventors or untrustworthy engineers. The typical recognizer gets it into his head that he can solve ‘the problem.’ The basis for this is either individual inspiration (the ‘mad inventor’ source of knowledge) or acceptance of untested rules, schemes, or information (the untrustworthy engineer approach). . . .”

“The typical recognizer ... builds or programs an elaborate system that either does very little or flops in an obscure way. A lot of money and time are spent. **No simple, clear, sure knowledge is gained.** The work has been an experience, not an experiment.”

Tell us what you really think, John

“We are safe in asserting that speech recognition is attractive to money. The attraction is perhaps similar to the attraction of schemes for turning water into gasoline, extracting gold from the sea, curing cancer, or going to the moon. One doesn’t attract thoughtlessly given dollars by means of schemes for cutting the cost of soap by 10%. **To sell suckers, one uses deceit and offers glamor.”**

“It is clear that glamor and any deceit in the field of speech recognition blind the takers of funds as much as they blind the givers of funds. Thus, we may pity workers whom we cannot respect.”

Fallout from these blasts

The first idea: **Try Artificial Intelligence . . .**

DARPA Speech Understanding Research Project (1972-75)

Used classical AI to try to “understand what is being said
with something of the facility of a native speaker”

DARPA SUR was viewed as a failure; funding was cut off after three years

The second idea: **Give Up.**

1975-1986: No U.S. research funding for MT or ASR

Pierce was far from the only person
with a jaundiced view of R&D investment
in the area of human language technology.

By the mid 1980s,
many informed American research managers
were equally skeptical about the prospects.

At the same time,
many people believed that HLT was needed
and in principle was feasible.

1985: Should DARPA restart HLT?

Charles Wayne -- DARPA program manager – has an idea.

He'll design a speech recognition research program that

- protects against “glamour and deceit”
 - because there is a well-defined, objective evaluation metric
 - applied by a neutral agent (NIST)
 - on shared data sets; and
- and ensures that “simple, clear, sure knowledge is gained”
 - because participants must reveal their methods
 - to the sponsor and to one another
 - at the time that the evaluation results are revealed

In 1986 America,

**no other sort of ASR program
could have been gotten large-scale government funding.**

NIST (Dave Pallett) 1985

"Performance Assessment of Automatic Speech Recognizers",
J. of Research of the National Bureau of Standards:

Definitive tests to fully characterize automatic speech recognizer or system performance cannot be specified at present. However, it is possible to design and conduct performance assessment tests that make use of widely available speech data bases, use test procedures similar to those used by others, and that are well documented. These tests provide valuable benchmark data and informative, though limited, predictive power. **By contrast, tests that make use of speech data bases that are not made available to others and for which the test procedures and results are poorly documented provide little objective information on system performance.**

“Common Task” structure

- A detailed “evaluation plan”
 - developed in consultation with researchers
 - and published as the first step in the project.
- Automatic evaluation software
 - written and maintained by NIST
 - and published at the start of the project.
- **Shared data:**
 - Training and “dev(elopment) test” data is published at start of project;
 - “eval(uation) test” data is withheld for periodic public evaluations

Not everyone liked it

Many Piercians were skeptical:

“You can’t turn water into gasoline,
no matter what you measure.”

Many researchers were disgruntled:

“It’s like being in first grade again --
you’re told exactly what to do,
and then you’re tested over and over .”

But it worked.

Why did it work?

1. The obvious: it allowed funding to start
(because the project was glamour-and-deceit-proof)
and to continue
(because funders could measure progress over time)
2. Less obvious: it allowed project-internal hill climbing
 - because the evaluation metrics were automatic
 - and the evaluation code was public*This obvious way of working was a new idea to many!*
... and researchers who had objected to be tested twice a year
began testing themselves every hour...
3. Even less obvious: it created a culture
(because researchers shared methods and results
on shared data with a common metric)

**Participation in this culture became so valuable
that many research groups joined without funding**

What else it did

The *common task method* created a positive feedback loop.

When everyone's program has to interpret the same ambiguous evidence, ambiguity resolution becomes a sort of gambling game, which rewards the use of statistical methods.

Given the nature of speech and language, statistical methods need the largest possible training set, which reinforces the value of shared data.

Iterated train-and-test cycles on this gambling game are addictive; they create “**simple, clear, sure knowledge**”, which motivates participation in the common-task culture.

The past 25 years

Variants of this method

have been applied to many other problems:

machine translation, speaker identification, language identification, parsing, sense disambiguation, information retrieval, information extraction, summarization, question answering, OCR, sentiment analysis, ... , etc.

The general experience:

1. Error rates decline by a fixed percentage each year,
to an asymptote depending on task and data quality
2. Progress usually comes from many small improvements;
a change of 1% can be a reason to break out the champagne.
3. Shared data plays a crucial role – and is re-used in unexpected ways.
4. Glamour and deceit have been avoided.

...and a self-sustaining process was started!



TIMIT



Scholar

About 12,000 results (0.03 sec)

My Citations

Articles

[CITATION] Darpa **Timit**: Acoustic-phonetic Continuous Speech Corps CD-ROM

JS Garofolo - 1993 - US Department of Commerce, ...

Cited by 543 - Related articles - PennText Services - Library Search

Legal documents

[CITATION] **TIMIT** acoustic-phonetic continuous speech corpus

JS Garofolo, LF Lamel, WM Fisher, JG Fiscus... - Linguistic Data Consortium, 1993

Cited by 257 - Related articles

Any time

Since 2012

Since 2011

Since 2008

Custom range...

Speech database development at MIT: **TIMIT** and beyond

V Zue, S Seneff... - Speech Communication, 1990 - Elsevier

Abstract Automatic speech recognition by computers can provide the most natural and efficient method of communication between humans and computers. While in recent years high performance speech recognition systems are beginning to emerge from research ...

Cited by 179 - Related articles - PennText Services - All 2 versions

Recent additions

☒ include patents

☒ include citations

[CITATION] Getting started with the DARPA **TIMIT** CD-ROM: An acoustic phonetic continuous speech database

JS Garofolo - National Institute of Standards and Technology (NIST), ..., 1988


Cited by 168 - Related articles

MMI training for continuous phoneme recognition on the **TIMIT** database

S Kapadia, V Valtchev... - Acoustics, Speech, and ..., 1993 - ieeexplore.ieee.org

Abstract Experiences with a phoneme recognition system for the **TIMIT** database which uses multiple mixture continuous-density monophone HMMs (hidden Markov models) trained using MMI (maximum mutual information) is reported. A comprehensive set of results are ...

Cited by 88 - Related articles - PennText Services - BL Direct - All 13 versions



Scholar
About 335 results (0.04 sec)

Articles
Legal documents
Any time
Since 2012
Since 2011
Since 2008
Custom range...
Recent additions
☒ include patents
☒ include citations

[Evaluation of TIMIT sentence list equivalency with adult cochlear implant recipients](#)
SE King, JB Firszt, RM Reeder... - Journal of the ..., 2012 - ingentaconnect.com
Background: Current measures used to determine sentence recognition abilities in cochlear implant recipients often include tests with one talker and one rate of speech. Performance with these measures may not accurately represent the speech recognition abilities of the ...
Cited by 1


[A Comparison of Network Architectures](#)
A Graves - Supervised Sequence Labelling with Recurrent Neural ..., 2012 - Springer
... The data for the experiments came from the **TIMIT** corpus (Garofolo et al., 1993) of prompted speech, collected by Texas Instruments. The utterances in **TIMIT** were chosen to be phonetically rich, and the speakers represent a wide variety of American dialects. ...

[CITATION] [Analysis of acoustic parameters for consonant voicing classification in clean and telephone speech](#)
SM Lee... - The Journal of the Acoustical ..., 2012 - Acoustical Society of America
Related articles - All 4 versions

[Significance of Prosody for Speaker, Language and Speech Recognition](#)
L Mary - Extraction and Representation of Prosody for Speaker, ..., 2012 - Springer
... 1.2 and 1.3, respectively [7]. The same utterance Don't carry an oily rag like that recorded through three different channels, available in Texas Instruments and Massachusetts Institute of Technology (**TIMIT**) database, is used for comparing the effect of channel variations. ...

[Voice intensity based gender classification by using Simpson's rule with SVM](#)
M Alsulaiman, Z Ali... - Systems, Signals and ..., 2012 - ieeexplore.ieee.org
... A promising classification rate of 98.27% is achieved. Index Terms— Voice Intensity, Simpson's Rule, SVM, **TIMIT**, Area under the curve. ... 2. SPEECH CORPUS The DARPA **TIMIT** Acoustic-Phonetic Continuous Speech Corpus (**TIMIT**) [10] is used to perform the experiments. ...

[Spectro-temporal modulation energy based mask for robust speaker identification](#)
TS Chi, TH Lin... - The Journal of the Acoustical Society of ..., 2012 - link.aip.org
... An algorithm which distinguishes speech from non-speech based on spectro-temporal modulation energies is proposed and evaluated in robust text-independent closed-set speaker identification simulations using the **TIMIT** and GRID corpora. ...



Scholar
About 6,080 results (0.05 sec)

Articles

Legal documents

Any time

Since 2012

Since 2011

Since 2008

Custom range...

Recent additions

☒ include patents
 ☒ include citations

[SWITCHBOARD: Telephone speech corpus for research and development](#)
[JJ Godfrey, EC Holliman...](#) - [Acoustics, Speech, and ...](#), 1992 - [ieeexplore.ieee.org](#)

Abstract **SWITCHBOARD** is a large multispeaker **corpus** of conversational speech and text which should be of interest to researchers in speaker authentication and large vocabulary speech recognition. About 2500 conversations by 500 speakers from around the US were ...
[Cited by 991](#) - [Related articles](#) - [All 5 versions](#)

[\[PDF\] Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus](#)
[S Greenberg, J Hollenback...](#) - [International Conference on Spoken ...](#), 1996 - [Citeseer](#)

ABSTRACT Models of speech recognition (by both human and machine) have traditionally assumed the phoneme to serve as the fundamental unit of phonetic and phonological analysis. However, phoneme-centric models have failed to provide a convincing ...
[Cited by 132](#) - [Related articles](#) - [View as HTML](#) - [All 18 versions](#)

[The effects of handset variability on speaker recognition performance: Experiments on the switchboard corpus](#)
[DA Reynolds](#) - ... [Speech, and Signal Processing](#), 1996. [ICASSP ...](#), 1996 - [ieeexplore.ieee.org](#)

Abstract This paper presents an empirical study of the effects of handset variability on text-independent speaker recognition performance using the **Switchboard corpus**. Handset variability occurs when training speech is collected using one type of handset, but a ...
[Cited by 81](#) - [Related articles](#) - [BL Direct](#) - [All 10 versions](#)

[Approaches to topic identification on the switchboard corpus](#)
[J McDonough, K Ng, P Jeanrenaud...](#) - ... [Speech, and Signal ...](#), 1994 - [ieeexplore.ieee.org](#)

Abstract Topic identification (TID) is the automatic classification of speech messages into one of a known set of possible topics. The TID task can be view as having three principal components: 1) event generation, 2) keyword event selection, and 3) topic modeling. ...
[Cited by 71](#) - [Related articles](#) - [BL Direct](#) - [All 5 versions](#)

[\[BOOK\] Dysfluency annotation stylebook for the switchboard corpus](#)
[MW Meteer, AA Taylor, R MacIntyre...](#) - 1995 - [cs.cmu.edu](#)

The **Switchboard Corpus** comprises telephone conversations between two individuals regarding a specific topic. Since it is conversational speech, it contains fragments of words, interruptions, incomplete sentences, fillers and discourse markers which require ...
[Cited by 91](#) - [Related articles](#) - [View as HTML](#) - [Library Search](#) - [All 15 versions](#)

Google

Scholar About 192 results (0.05 sec)

Articles

Legal documents

Any time

Since 2012

Since 2011

Since 2008

Custom range...

Recent additions

☒ include patents

☒ include citations

[\[PDF\] Collaborative Annotation of Dialogue Acts: Application of a New ISO Standard to the Switchboard Corpus](#)
AC Fang, H Bunt, J Cao... - EACL 2012, 2012 - newdesign.aclweb.org
Abstract This article reports some initial results from the collaborative work on converting SWBD-DAMSL annotation scheme used in the **Switchboard** Dialogue Act **Corpus** to ISO DA annotation framework, as part of our on-going research on the interoperability of ...
[View as HTML](#)

[\[PDF\] Conceptual effects on agreement: A corpus study of mismatch in English copular constructions](#)
CJ Duffield - LSA Annual Meeting Extended Abstracts, 2012 - elanguage.net
... Examples of singular (SG) and plural (PL) subject-verb agreement are shown in utterances (1-2) taken from the **Switchboard Corpus** (Godfrey, Holliman & McDaniel, 1992). ...
SWITCHBOARD: Telephone Speech **Corpus** for Research and Development. ...
[View as HTML](#)

[\[PDF\] Automatic Animacy Classification](#)
SR Bowman... - NAACL-HLT 2012, 2012 - aclweb.org
... The **corpus** that we use is Zaenen et al.'s animacy- annotated subset of the hand-parsed **Switchboard corpus** of conversational American English. It is built on, and now included in, Calhoun et al.'s (2010) NXT version of **Switchboard**. ... 2010. The NXT- format **Switchboard Corpus**. ...
[View as HTML](#)

[Estimating the number of segments for improving dialogue act labelling](#)
V Tamarit... - Natural Language ..., 2012 - Cambridge Univ Press
... The **SwitchBoard corpus** is a well-known **corpus** in English. It is composed of recorded conversations between humans with no particular goal to accomplish, so it does not represent a real dialogue system. ... 4.1 **SwitchBoard Corpus** The **SwitchBoard corpus** (Godfrey et al. ...
Cited by 1 - Related articles - All 3 versions

[\[PDF\] Presentational unaccusativity: Argument structure and information status](#)
P Irwin - 2012 - files.nyu.edu
... Francis et al. (1999): distribution of subject and object forms in the **Switchboard Corpus** • pronominal: 91% of subjects ... 5.2 Method 5.2.1 **Corpus** and sentence extraction **Switchboard Corpus** • Subset of the **corpus** annotated for DP information status (Godfrey et al., 1992; ...
Related articles - View as HTML

Automatic Content Extraction (ACE) Evaluation

What is Automatic Content Extraction (ACE)?

The objective of the ACE program is to develop automatic content extraction technology to support automatic processing of human language in text form from a variety of sources (such as newswire, broadcast conversation, and weblogs). ACE technology R&D is aimed at supporting various classification filtering, and selection applications by extracting and representing language content (i.e., the meaning conveyed by the data). Thus the ACE program requires the development of technologies that automatically detect and characterize this meaning.

The ACE program will be carried out in several phases, beginning with EDT (Entity Detection and Tracking) Phase-1.

View the [presentation](#) describing the ACE program that was given at the TIDES program kick-off meeting.

Current and Recent ACE Activities

The most recent ACE evaluation was [ACE08](#) and took place in May 2008.

Results of recent ACE Evaluations:

- [NIST ACE08 Official Evaluation Results](#)
- [NIST ACE07 Official Evaluation results](#)

ACE is becoming a track in the Text Analysis Conference (TAC) in 2009. Please explore the [TAC website](#).

Find more information on past ACE evaluations by clicking a specific year in the tabs below.

[[1999](#)] [[2000](#)] [[2001](#)] [[2002](#)] [[2003](#)] [[2004](#)] [[2005](#)] [[2007](#)] [[2008](#)]



[About TAC](#)
[TAC KBP 2012](#)
[WEAS 2012](#)
[Past Tracks](#)
[Past Data](#)
[Publications](#)
[Contact](#)



TAC Tracks

A TAC cycle consists of a set tracks -- areas of focus in which particular NLP tasks are defined. Task definitions and guidelines for a particular year are given in the individual track home pages for that year (see the table of track home pages below). Data for each track in the current TAC cycle (including test data and evaluation results) are available only to registered participants in the current cycle. Data from past TACs are available to non-participants on the web page for [Past Data](#).

TAC Track Home Pages				
All Tracks	2008	2009	2010	2011
Question Answering	2008			
Recognizing Textual Entailment	2008	2009	2010	2011
Summarization	2008	2009	2010	2011
Knowledge Base Population		2009	2010	2011

Mailing Lists

Each track has a mailing list. The primary purpose of the mailing list is to discuss the details of the track's tasks in the current TAC cycle. However, a track mailing list also serves as a place to discuss general methodological issues related to the track's NLP focus. TAC track mailing lists are open to all; you need not participate in TAC to join a list. Most lists do require that you become a member of the list before you can send a message to it. Instructions for joining track mailing lists are given in the individual track home pages.



About TAC
TAC KBP 2012
WEAS 2012
Past Tracks
Past Data
Publications
Contact



TAC KBP 2012

- [Call for Participation](#)
- [Registration](#)
- [Task Guidelines](#)
- [Data](#)
- [Tools](#)
- [TAC 2012 Workshop](#)

TAC KBP 2012 Overview

TAC 2012 focuses on Knowledge Base Population (KBP). The goal of Knowledge Base Population is to promote research in automated systems that discover information about named entities as found in a large corpus and incorporate this information into a knowledge base. TAC 2012 fields tasks in three areas, all aimed at improving the ability to automatically populate knowledge bases from text:

1. **Entity-Linking:** Given a name (of a Person, Organization, or Geopolitical Entity) and a document containing that name, determine the KB node for the named entity, adding a new node for the entity if it is not already in the KB. The reference KB is derived from English Wikipedia, while source documents come from a variety of languages, including English, Chinese, and Spanish.
2. **Slot-Filling:** Given a named entity and a pre-defined set of attributes ("slots") for the entity type, augment a KB node for that entity by extracting all new learnable slot values for the entity as found in a large corpus of documents. The reference KB is derived from English Wikipedia, while source documents come from English and Spanish. A diagnostic task, Slot Filler Validation, will be to determine whether a candidate filler in a document is a correct slot-filler for a given entity.
3. **Cold Start Knowledge Base Population:** Given a KB schema with an empty knowledge base, build the KB from scratch by mining a large text collection.

To promote research in populating probabilistic knowledge bases, systems may augment each assertion they make with a confidence score.

[\[PDF\] The automatic content extraction \(ace\) program—tasks, data, and evaluation](#)[\[PDF\] from psu.edu](#)

G Doddington, A Mitchell, M Przybocki... - Proceedings of ..., 2004 - Citeseer

... References • LDC, 2004, **Automatic Content Extraction** [www ldc.upenn.edu/Projects/ACE/] •

NIST, 1999, Message Understanding Conference [www.itl.nist.gov/iaui/894.02/related_projects/

muc/] • NIST, 2003, **Automatic Content Extraction** [www.nist.gov/speech/tests/ace/] • NIST ...[Cited by 115](#) - [Related articles](#) - [View as HTML](#) - [All 8 versions](#)[\[PDF\] Linguistic resources and evaluation techniques for evaluation of cross-document automatic content extraction](#)[\[PDF\] from unipi.it](#)

S Strassel, M Przybocki, K Peterson... - Proceedings of the ..., 2008 - mailserver.di.unipi.it

... Abstract The **NIST Automatic Content Extraction (ACE)** Evaluation expands its focus

in 2008 to encompass the challenge of cross-document and cross-language global

integration and reconciliation of information. While past ...

[Cited by 13](#) - [Related articles](#) - [View as HTML](#) - [All 8 versions](#)[Adapting a robust multi-genre NE system for automatic content extraction](#)

D Maynard, H Cunningham, K Bontcheva... - Artificial Intelligence: ..., 2002 - Springer

... 2. K. Bontcheva, D. Maynard, H. Saggion, and H. Cunningham. Using human lan-

guage technology for **automatic** annotation and indexing of digital library **content**. ...

Nist's 1998 topic detection and tracking evaluation (tdt2). In Proc. ...

[Cited by 16](#) - [Related articles](#) - [BL Direct](#) - [All 9 versions](#)[\[PDF\] Information extraction, automatic](#)[\[PDF\] from psu.edu](#)

H Cunningham - Encyclopedia of Language and Linguistics, 2005 - Citeseer

... The **Automatic Content Extraction** programme (ACE, ACE (2004); Maynard et al. ... It annotateswebpages with metadata in a fully **automatic** fashion and needs no manual intervention. ... MessageUnderstanding Conference (MUC-7). Fairfax, VA, p. 5 pages, <http://www.itl.nist.gov/iaui> ...[Cited by 87](#) - [Related articles](#) - [View as HTML](#) - [All 4 versions](#)[Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations](#)[\[PDF\] from upenn.edu](#)

N Kambhatla - Proceedings of the ACL 2004 on Interactive poster ..., 2004 - portal.acm.org

... In this paper, we present our general approach, describe the features we currently use and

show the results of our participation in the ACE evaluation. **Automatic Content Extraction**(ACE, 2004) is an evaluation conducted by **NIST** to measure Entity ...[Cited by 135](#) - [Related articles](#) - [All 11 versions](#)

The Linguistic Data Consortium

- Started in 1992 with seed money from DARPA
- LDC has distributed
 - more than 84,000 copies
 - of more than 1,300 titles
 - to more than 3,168 organizations
 - in more than 70 countries
- About half of the titles are “common task” datasets
 - developed for technology evaluation programs
 - released via general catalog after program use
- ~30 titles added to general catalog per year
(current publications queue of about 120 items)

Where we were

ANLP-1983

(First Conference on Applied Natural Language Processing)

34 Presentations:

None use a published data set.

None use a formal evaluation metric.

Two examples:

Wendy Lehnert and Steven Shwartz,

“EXPLORER: A Natural Language Processing System for Oil Exploration”.

Describes problem and system architecture; gives examples of queries and responses.

No way to evaluate performance or to compare to other systems/approaches.

Larry Reeker et al.,

“Specialized Information Extraction: Automatic Chemical Reaction Coding from English Descriptions”

Describes problem and system architecture; gives examples of inputs and outputs.

No way to evaluate performance or to compare to other systems/approaches.

Where we are

ACL-2010

(48th Annual Meeting of the Association for Computational Linguistics)

274 presentations –

Essentially all use published data and published evaluation methods.
(A few deal with new data-set creation and/or new evaluation metrics.)

Three examples:

Nils Reiter and Anette Frank, "Identifying Generic Noun Phrases".
Authors are from Heidelberg University; use ACE-2 data.

Shih-Hsiang Lin and Berlin Chen,
"A Risk Minimization Framework for Extractive Speech Summarization".
Authors are from National Taiwan University;
use Academia Sinica Broadcast News Corpus
and the ROUGE metric (developed in DUC summarization track).

Laura Chiticariu et al., "An Algebraic Approach to Declarative Information Extraction".
Authors are from IBM Research; use ACE NER metric, ACE data, ENRON corpus data.

Science is different...

But not that different:

Sharing data and problems

- lowers costs and barriers to entry
- creates intellectual communities
- speeds up replication and extension
- and guards against “glamour and deceit”
(...as well as simple confusion)

Thank you!

1981: Doddington, IEEE Spectrum

Doddington and Schalk,

"Speech Recognition: Turning Theory into Practice"

[Documents a formal evaluation of 7 commercially-available isolated-word recognizers, using training and testing data from 16 speakers.]

"TI's Central Research Laboratory has been called upon frequently to consult with other groups at TI on the use of word recognition, to make recommendations regarding the purchase of word recognizers, and to project system performance. It was eventually concluded that the only responsible way to perform these tasks was to formally evaluate the performance of candidate systems."