UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ROKU, INC.
Petitioner

v.
UNIVERSAL ELECTRONICS, INC.
Patent Owner.

EXPERT DECLARATION OF DR. GRETCHEN HOFFMAN IN SUPPORT OF ROKU, INC.'S PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,847,083

I, Gretchen Hoffman, declare as follows:

1. I am presently a Professor in the School of Library and Information Studies at Texas Woman's University in Denton, Texas.

2. I have prepared this declaration in connection with Roku, Inc.'s Petition for Inter Partes Review of U.S. Patent No. 9,847,083 ("the '083 Patent"), which is to be filed concurrently with this declaration.

3. This declaration sets forth my opinions based on the information identified and the analyses I have performed through the date of this declaration. This declaration, and my opinions contained herein, are subject to change or modification if additional relevant information becomes available that bears on my analysis.

4. I have been retained by Roku, Inc. ("Roku" or "Petitioner") as an expert in the field of library cataloging. I am being compensated at my normal consulting rate of $350 per hour for my time. My compensation does not depend in any way on the substance of my conclusions and is not affected by the outcome of this proceeding.

I. EXPERIENCE AND QUALIFICATIONS

5. I received a Bachelor of Arts degree in Music from the University of Nevada, Reno in 1995 and a Master of Library Science degree from Emporia State University in 2000. I received a doctoral degree in Library and Information Management from Emporia State University in 2008. I have been a professional

librarian for more than 20 years and I am recognized as a leading authority in library cataloging and information management, both as a professor and as a practitioner.

6. I am a long-standing member of state, regional, and national professional organizations. I am a member of the American Library Association ("ALA") and its Core: Leadership, Infrastructure, Futures division, Library Research Round Table ("LRRT"), and Competencies and Education for a Career in Cataloging Interest Group. I am also a member of the Association for Information Science and Technology ("ASIS&T"), the Association for Library and Information Science Education ("ALISE"), the International Society for Knowledge Organization ("ISKO"), and the Texas Library Association ("TLA"). I served as the Chair of the Cataloging and Metadata Round Table of the Texas Library Association between 2011 and 2012. I have also served as a peer reviewer and subject matter expert for several publications within the library and information science field, including *Library Resources & Technical Services, Journal of Library Metadata, and Cataloging & Classification Quarterly*.

7. I published a textbook in fall 2019, *Organizing Library Collections: Theory and Practice*, that focuses on the theory and practice of organizing library collections. I have given numerous presentations in the field of library cataloging and information management, including presentations on the topics of library cataloging and classification systems, bibliographic control, cataloging ethics, and metadata. I have

also conducted numerous field-based research studies, and published numerous book chapters and articles.

8. Additional details concerning my professional qualifications, experience, and publications are set forth in my current curriculum vitae, attached hereto as Attachment Z.

II. RELEVANT LEGAL STANDARDS

9. I am not an attorney and have not been asked to offer my opinion on the law. As an expert offering an opinion on whether documents qualify as printed publications, however, I understand that I am obliged to follow existing law. For the purposes of this Declaration, I have been informed about certain aspects of the law that are relevant to my opinions. My understanding of the law is summarized below.

10. I have been informed and understand that under 35 U.S.C. § 102, certain materials may qualify as prior art in patent litigation proceedings when that material constitutes a "printed publication." I understand that to qualify as a printed publication, the material in question must be sufficiently accessible to the public interested in the art. In turn, material is sufficiently accessible if it was disseminated or otherwise made available to the extent that persons interested and ordinarily skilled in the subject matter or art exercising reasonable diligence, can locate it. I further understand that accessibility goes to the issue of whether interested members of the relevant public could obtain the information if they wanted to, and not whether any particular member of the public actually retrieved the material.

11. I understand that one way to demonstrate public accessibility is to show competent evidence of general library practices indicating an approximate date on which the material in question became publicly accessible. For example, I understand that cataloging or indexing material in a library can show the material was sufficiently accessible to those interested in the art. I also understand that this is not the only way to prove public accessibility. Among other ways, public accessibility may be proven by the copyright date of the material in question or by citations to the material in question that are contained in other materials.

III. LIBRARY CATALOGING

12. OCLC, formerly named the Online Computer Library Center, is a worldwide library cooperative that is owned and governed by its participating members, and one of its goals is to establish, maintain, and operate a computerized library network. Over 23,000 libraries, archives, and museums in over 170 countries (including many libraries in the United States) are members of OCLC.

13. When an OCLC participating institution acquires a work, a cataloger creates a bibliographic record for this work using the Machine-Readable Cataloging ("MARC") format in its computer catalog system and OCLC automatically supplies the date of creation. The MARC format is an industry-wide standard method of encoding, storing, and organizing library catalog information. MARC is the primary communications protocol for the transfer and storage of bibliographic metadata in libraries.

14. The creation date of a bibliographic record encoded in the MARC format reflects the date on or shortly after the item was first acquired and cataloged. Once the bibliographic record is created by a cataloger at an OCLC participating member institution, it is instantaneously available to any OCLC participating members, and therefore available to the public. Many bibliographic records on the OCLC bibliographic database are created originally by the Library of Congress and other national libraries, such as the British Library. The general public has free online access to records created by OCLC member libraries in OCLC's WorldCat database.

15. A bibliographic record encoded in the MARC format includes several fields that contain specific data about the work. For example, a work's title and authors or editors are recorded in field 245. The edition statement is recorded in field 250. The publisher and date appear in field 260 or 264. An individual determined to be the primary creator of a work appears in field 100; secondary creators, editors, and other contributors appear in field 700. If a corporate body is determined to be responsible for creating the intellectual content of a work in lieu of one or more individuals, then its name appears in field 110 (when a creator) or field 710 (when a contributor). Notes are added in 5XX fields. 9XX fields are local fields that can be used and defined by each library. They usually list information about when the item was ordered, received, cataloged, and/or barcoded by the library. Field 008 records, among other things, the date a bibliographic record was created. Field 005 records the date a bibliographic record was last updated.

16. Bibliographic records are indexed in several ways in library catalogs. Titles are indexed so users may perform keyword searches or direct searches for titles, including titles, series titles, and variant titles. Names are indexed so users may perform keyword searches or direct searches for names, including personal names, corporate names, and family names. Subjects are indexed by adding subject headings and classification numbers based on subject. Subjects are indexed by using various subject headings lists and classification schemes maintained by the library profession, such as the *Library of Congress Subject Headings* and the *Library of Congress Classification* scheme. In addition, users are able to search bibliographic records using keyword searches, which allow users to search the entire bibliographic record, including titles, names, subjects, and when added to bibliographic records, table of contents and notes.

17. In addition to books that are cataloged and indexed through the online library catalog, libraries also provide access to journal articles and other serial publications (e.g., journals, magazines, newspapers, newsletters, etc.). A library can subscribe to the print version of a serial publication that is mailed to the library, and/or a library can subscribe to the online version available directly through a publisher or through reference databases. Of note in this declaration is that some serial publications are open access, meaning they are publicly available online for free, and libraries routinely provide links to open access serials publications through the online library catalog. In addition, many libraries provide access to digital collections, archival materials, and special collections. Finally, libraries provide their patrons access to the

World Wide Web through computers available in the library, thus extending their services to patrons and increasing the number of materials they can provide to patrons.

18. Authoritative databases, such as the ScienceDirect database, the Web of Science database, the IEEE Xplore database, the SpringerOpen database, DOAJ: Directory of Open Access Journals database, the Internet Archive's Wayback Machine, and online library catalogs for the Library of Congress, the United States Copyright Office, the British Library, and several university libraries were used to confirm citation details of the references identified in this declaration.

IV. "SPEECH RECOGNITION, AUTOMATIC: HISTORY" (EXHIBIT 1012)

19. I understand that the document "Speech Recognition, Automatic: History" is to be submitted concurrently with this Declaration as Exhibit 1012.

20. Based on my review of Exhibit 1012, it appears to be an encyclopedia article titled "Speech Recognition, Automatic: History" written by B.-H. Juang and L.R. Rabiner, published by Elsevier, with a copyright date of 2006 (hereafter "Juang and Rabiner article"). Exhibit 1012 is a copy of the Juang and Rabiner article, which is located on pages 806-819 of the encyclopedia. I received a copy of this article from counsel, who, I understand, bought the copy which comprises Exhibit 1012 directly from the publisher. I note that those parts of the text that comprise Exhibit 1012 appear to be complete and there appear to be no visible alterations to the document.

21. To determine when the Juang and Rabiner article was cataloged and made available in libraries, it is first necessary to find the title of the encyclopedia in which

the article was published. Attached hereto as Attachment A is a true and correct copy of a database record from the ScienceDirect database that shows where the article was published. ScienceDirect is an authoritative database from Elsevier that indexes scientific, heath, and technical publications. The citation information in the database record matches the citation information of Exhibit 1012. The database record indicates that "Speech Recognition, Automatic: History" by B.-H. Juang and L.R. Rabiner, was published in 2006 on pages 806-819. It further explains that the article was published in the *Encyclopedia of Language & Linguistics*, second edition.

22. Libraries catalog encyclopedias as a whole work; they generally do not catalog each volume separately, and they do not index each individual article separately (that indexing is done by reference databases available in libraries, such as ScienceDirect). To determine when the second edition of *Encyclopedia of Language & Linguistics*, including the Juang and Rabiner article, was made publicly available, it is helpful to look at bibliographic records from online library catalogs. Attached hereto as Attachment B is a true and correct copy of the MARC bibliographic record for the second edition of *Encyclopedia of Language & Linguistics* that I obtained from the shared online library catalog of Indiana University (IU) that includes holdings for all universities in the IU system. The bibliographic identifiers match those of Attachment A, including title (field 245), edition (field 250), and date (field 260). As shown in field 300, the second edition of *Encyclopedia of Language & Linguistics* was published in 14 volumes. Field 008 on the bibliographic record shows that the bibliographic record was

created on April 28, 2005, downloaded from the OCLC bibliographic database, as indicated in the second field 035. Field 005 explains when the bibliographic record in the shared online library catalog of Indiana University was last updated, which was March 3, 2007. The multiple 999 fields on this bibliographic show the barcode for each volume of the encyclopedia and which campus libraries hold them. The first field 999 shows that the Indiana University Bloomington Library ("B-WELLS") barcoded each volume of the encyclopedia on August 23, 2006. The 15th field 999 shows that Indiana University South Bend Library ("SBEND") barcoded each volume of the encyclopedia on June 20, 2007. These dates are an indication when the second edition of *Encyclopedia of Language & Linguistics* was cataloged and made available at the Indiana University Bloomington Library and the Indiana University South Bend Library. It is my opinion that the second edition of *Encyclopedia of Language & Linguistics*, including the Juang and Rabiner article, was published and accessible to the public at least as early as June 20, 2007.

23. My opinion is further substantiated by a bibliographic record in another library that holds the second edition of *Encyclopedia of Language & Linguistics*. Attached hereto as Attachment C is a true and correct copy of the MARC bibliographic record for the second edition of *Encyclopedia of Language & Linguistics* from the online library catalog of the University of Michigan that I obtained. The bibliographic identifiers match those of Attachment A and Attachment B, including title (field 245), edition (field 250), date (field 260), and number of volumes (field 300). The multiple

974 fields on this bibliographic show when each volume of the encyclopedia was barcoded and made available. The dates range from December 20, 2005 to February 1, 2006. These dates are an indication when the second edition of *Encyclopedia of Language & Linguistics* was cataloged and made available at the University of Michigan Library. This is around 6 months before the encyclopedia was made available at the Indiana University Bloomington Library and over a year before the encyclopedia was made available at the Indiana University South Bend Library.

24. Furthermore, the database record in Attachment A includes 15 indexing descriptors that provide subject access to the Juang and Rabiner article: "acoustic modeling; automatic transcription; dialog systems; finite state network; hidden Markov models; keyword spotting; language modeling; neural networks; office automation; pattern recognition; spectral analysis; Speech recognition; speech understanding; statistical modeling; time normalization." There is subject access to the second edition of *Encyclopedia of Language & Linguistics* as well. The bibliographic records in Attachments B and C include two English-language subject headings in field 650 reading "Language and languages |v Encyclopedias" and "Linguistics |v Encyclopedias". The "_0" in field 650 indicates the subject headings were taken from the *Library of Congress Subject Headings*. In addition, Attachment B includes a *Library of Congress Classification* number based on subject in field 090 and Attachment C includes a *Library of Congress Classification* number based on subject in field 050 that read "P29 .E48 2006". The first part of this classification number ("P29")

represents the subject “Philology. Linguistics" and is assigned to dictionaries and encyclopedias about the subject. Thus, the Juang and Rabiner article, published in the second edition of *Encyclopedia of Language & Linguistics*—as shown in the database record in Attachment A and the MARC bibliographic records in Attachments B and C—was indexed according to its subject matter.

25. Based on my review of Exhibit 1012, my review of the database record in the ScienceDirect database, and the bibliographic records in the shared online library catalog of the Indiana University system and the online library catalog of the University of Michigan, as well as my experience in library cataloging, it is my opinion that the Juang and Rabiner article, published in the second edition of *Encyclopedia of Language & Linguistics*, was publicly accessible at least as early as June 20, 2007. Accordingly, it is my opinion that the Juang and Rabiner article was accessible to the public prior to Nov. 17, 2011.

V. “HISTORY OF SPEECH & VOICE RECOGNITION AND TRANSCRIPTION SOFTWARE” (EXHIBIT 1013)

26. I understand that a document titled “History of Speech & Voice Recognition and Transcription Software” is to be submitted concurrently with this Declaration as Exhibit 1013.

27. Based on my review of Exhibit 1013, it appears on its face to be a webpage titled “History of Speech & Voice Recognition and Transcription Software” (hereafter “HSVRTS webpage”), and there is no author cited on the webpage. It appears to be

from the Dragon Naturally Speaking Medical Transcription website, given that the URL of "http://www.dragon-medical-transcription.com" appears on the top of the first page of Exhibit 1013, and "Dragon Naturally Speaking Medical Transcription Software" is at the top of the sidebar of the left-hand side of the webpage. I received a copy of this webpage from counsel, who, I understand, obtained Exhibit 1013 from the Internet Archive's Wayback Machine. I note that those parts of the text that comprise Exhibit 1013 appear to be complete and there appear to be no visible alterations to the document.

28. The first page of Exhibit 1013 appears to show that the printout was captured by the Internet Archive's Wayback Machine on March 19, 2005. This is supported by the URL on the second page of the exhibit, which includes the March 19, 2005 date as part of the URL ("20050319"). To confirm this information, I reviewed the web-based archives of Exhibit 1013 on the Internet Archive's Wayback Machine, a widely-used database of web-based archives. Wayback Machine is a publicly available service, available at https://archive.org/web, for searching and viewing archived digital internet webpages. Wayback Machine uses software that surfs the Internet and automatically captures and stores copies of publicly available webpages, preserving them as they existed at the time of capture. A captured webpage can be retrieved from the Wayback Machine archives using the corresponding Uniform Resource Locator (URL) of the webpage. I have used the Wayback Machine service to search the

availability of publications on the web and consider it to be a reliable and unbiased source.

29. I entered the URL on the top margin of first page of Exhibit 1013 in the search field of Wayback Machine. Attached hereto as Attachment D is a true and correct copy of my search for the HSVRTS webpage on the Wayback Machine. It is the same webpage as Exhibit 1013, and includes the same title, URL, and content. Attachment D indicates that this URL was first captured on March 19, 2005, and it has been captured by the Wayback Machine 261 times since then. It is my opinion that the HSVRTS webpage was accessible to the public on the "http://www.dragon-medical-transcription.com" website at least as early as March 19, 2005.

30. My opinion is supported by the Affidavit of Nathaniel E. Frank-White, a Records Request Processor at the Internet Archive, which was submitted separately as Exhibit 1041. In his Attachment A, Mr. Frank-White confirms that Exhibit 1013 was first captured by the Internet Archive's Wayback Machine on March, 19, 2005. The title, URL, and content in Mr. Frank-White's Appendix A match the title, URL, and content in Exhibit 1013 and this Declaration's Attachment D.

31. Based on my review of Exhibit 1013, the web-based archives, the Affidavit of Nathaniel E. Frank-White, as well as my experience in library cataloging, it is my opinion that the contents of Exhibit 1013 were available on the "http://www.dragon-medical-transcription.com" website at least as early as March 19, 2005. Accordingly, it

is my opinion that the HSVRTS webpage was accessible to the public prior to Nov. 17, 2011.

VI. "AN OVERVIEW OF MODERN SPEECH RECOGNITION" (EXHIBIT 1014)

32. I understand a document titled "An Overview of Modern Speech Recognition" is to be submitted concurrently with this Declaration as Exhibit 1014.

33. Based on my review, Exhibit 1014 appears to be a book chapter titled "An Overview of Modern Speech Recognition" by Xuedong Huang and Li Deng (hereafter "Huang and Deng chapter"). As indicated by the title page and contents page of Exhibit 1014, the Huang and Deng chapter was published as chapter 15 in the edited collection *Handbook of Natural Language Processing*, second edition, edited by Nitin Indurkhya and Fred J. Damerau, and published by Chapman & Hall/CRC. Exhibit 1014 is a copy of parts of the second edition of *Handbook of Natural Language Processing*, including the front cover, back cover, spine, title page and verso, contents, and the Huang and Deng chapter on pages 339-366, made from a print copy obtained from the Library of Congress. The handbook's title page verso shows a copyright date of 2010, shows a 13-digit International Standard Book Number (ISBN): 978-1-4200-8592-1, shows a call number: QA76.9.N38 H363 2010, and shows a Library of Congress Control Number (LCCN): 2009049507. The verso also bears a handwritten call number at the Library of Congress: QA76.9.N38 H363 2010 Copy 2. The spine bears the same call number label: QA76.9.N38 H363 2010 Copy 2. The back cover bears a barcode label: LC Copyright

0 030 853 937 4. I received a copy of Exhibit 1014 from counsel, who, I understand, obtained the copy from the Library of Congress. I note that those parts of the text that comprise Exhibit 1014 appear to be complete and there appear to be no visible alterations to the document.

34. Attached hereto as Attachment E is a true and correct copy of the online catalog record for the second edition of *Handbook of Natural Language Processing* from the Library of Congress online catalog that I obtained. The document cataloged in this record is the second edition of *Handbook of Natural Language Processing* as verified by the fields listing the title: *Handbook of Natural Language Processing*; publisher and copyright date: Chapman & Hall/CRC in 2010, 13-digit ISBN: 9781420085921 and LCCN: 2009049507. I note that the record does not include "second edition". However, in my experience, this is an accidental omission by the Library of Congress because everything else on the MARC bibliographic record corresponds to the second edition of *Handbook of Natural Language Processing*. I also compared the LC Classification/Call Number (QA76.9.N38 H363 2010) with that shown on the spine and verso of Exhibit 1014, and the values are equivalent on both. The record indicates that two print copies of the book are held by the Library of Congress. The Attachment indicates that the second edition of *Handbook of Natural Language Processing*, including the Huang and Deng chapter, as cataloged at the Library of Congress is currently held by that institution. All identifying characteristics, including stamps and notations, on Exhibit 1014 are consistent with library practice and

procedure that I have observed during my career as a professional librarian. Based on my review of Exhibit 1014, in my opinion, Exhibit 1014 is a true and correct copy in a condition that creates no suspicion about its authenticity.

35. Libraries catalog edited collections as a whole work; they generally do not catalog each book chapter separately, and they do not index each book chapter separately (that indexing is done by reference databases available in libraries, such as Web of Science). To determine the date that the second edition of *Handbook of Natural Language Processing*, including the Huang and Deng chapter, was made publicly available in libraries, I looked at the MARC bibliographic record in several libraries that hold the second edition of *Handbook of Natural Language Processing.* Attached hereto as Attachment F is a true and correct copy of the MARC bibliographic record for the second edition of *Handbook of Natural Language Processing* that I obtained from the shared online library catalog of the Indiana University system. The bibliographic identifiers of the second edition of *Handbook of Natural Language Processing* match those of Exhibit 1014, including the including title and editors (field 245), publisher and copyright date (field 260), 13-digit ISBN (field 020), Library of Congress Control Number (field 010), Library of Congress Classification number (field 050), and so on. In addition, this bibliographic record includes field 520, which confirms the "Second Edition" of the handbook, and field 505, which confirms that the 15th chapter of the handbook is "An overview of modern speech recognition / Xuedong Huang and Li Deng", therefore confirming that the bibliographic record corresponds to Exhibit 1014.

Field 008 on the bibliographic record shows that the bibliographic record was created on November 23, 2009, downloaded from the OCLC bibliographic database, as indicated in the second and third field 035. Field 005 explains when the bibliographic record in the shared online library catalog of Indiana University was last updated, which was October 14, 2010. The two 999 fields on this bibliographic record show the barcode for each volume of the handbook and which campus libraries hold them. The first field 999 shows that the Indiana University-Purdue University Indianapolis Library ("I-UNIVLIB") barcoded the handbook on May 12, 2011. The second field 999 shows that the Indiana University South Bend Library ("SBEND") barcoded the handbook on April 4, 2011. These dates are an indication when the second edition of *Handbook of Natural Language Processing* was cataloged and made available at the Indiana University-Purdue University Indianapolis (IUPUI) Library and the Indiana University South Bend Library. It is my opinion that the second edition of *Handbook of Natural Language Processing*, including the Huang and Deng chapter, was published and accessible to the public at least as early as May 12, 2011.

36. My opinion is further substantiated by bibliographic records from other libraries that hold the second edition of *Handbook of Natural Language Processing*. Attached hereto as Attachment G is a true and correct copy of the MARC bibliographic record for the second edition of *Handbook of Natural Language Processing* from the online library catalog of Washington University in St. Louis that I obtained. The bibliographic identifiers match those of Attachment F, including title and editors (field

245), publisher and copyright date (field 260), 13-digit ISBN (field 020), Library of Congress Control Number (010), Library of Congress Classification number (field 050), and so on. In addition, this bibliographic record includes field 520, which confirms the "Second Edition" of the handbook, and the 19th field 970 shows that the Huang and Deng chapter is the 15th chapter of the handbook. Field 008 on the bibliographic record shows that the bibliographic record was created on November 23, 2009, downloaded from the OCLC bibliographic database, as indicated in field 035. Field 005 explains when the bibliographic record in the shared online library catalog of Washington University in St. Louis was last updated, which was April 16, 2010. This date is an indication when the second edition of *Handbook of Natural Language Processing* was cataloged at the Washington University in St. Louis Library, although it may have taken several months for the item to be barcoded and shelved. This is around one year or so before the handbook was made available at the Indiana University-Purdue University Indianapolis (IUPUI) Library and the Indiana University South Bend Library.

37. Attached hereto as Attachment H is a true and correct copy of the MARC bibliographic record for the second edition of *Handbook of Natural Language Processing*, including the Huang and Deng chapter, that I obtained from the online library catalog of the Library of Congress. The bibliographic identifiers match those of Attachments F and G, including title and editors (field 245), publisher and copyright date (field 260), 13-digit ISBN (field 020), Library of Congress Control Number (field 010), Library of Congress Classification number (field 050), and so on. I note that the

MARC bibliographic record does not mention the "Second Edition" like on the MARC bibliographic records in Attachments F and G. However, in my experience, this is just an accidental omission by the Library of Congress because everything else on the MARC bibliographic record corresponds to the second edition of *Handbook of Natural Language Processing*. Field 008 on the bibliographic record shows that the bibliographic record was created on November 23, 2009. Field 005 explains when the bibliographic record in the Library of Congress's online catalog was last updated, which was June 2, 2010. The 955 field on this record shows when the handbook moved from different people/departments at the Library of Congress. The latest date is June 2, 2010, and the information indicates when a second copy of the handbook was added to the library's collection ("2010-06-02 c. 2 added"). Because a second copy of the second edition of *Handbook of Natural Language Processing*, including the Huang and Deng chapter, was in the library on June 2, 2010, and the bibliographic record has not been updated since June 2, 2010, one can reasonably assume the handbook was available on that date. This is around 10-11 months before the handbook was made available at the Indiana University-Purdue University Indianapolis (IUPUI) Library and the Indiana University South Bend Library.

38. Furthermore, the Huang and Deng chapter and the *Handbook of Natural Language Processing* were indexed according to their subject matter. Attached hereto as Attachment I is a true and correct copy of a database record for the Huang and Deng chapter from the Web of Science database. Web of Science is an authoritative database

from Clarivate that indexes publications in many subject areas. The citation information in the database record matches the citation information of Exhibit 1014. The database record includes 3 indexing descriptors ("Keywords Plus") that provide subject access to the Huang and Deng chapter: MAXIMUM-LIKELIHOOD; REGRESSION; HMMS. There is subject access to the second edition of *Handbook of Natural Language Processing* as well. The MARC bibliographic records in Attachments F, G, and H include one subject heading in field 650 reading "Natural language processing (Computer science) |v Handbooks, manuals, etc." The "_0" in field 650 indicates the subject heading was taken from the *Library of Congress Subject Headings*. In addition, Attachments F, G, and H include a *Library of Congress Classification* number based on subject in field 050 ("QA76.9.N38 H363 2010"). The first part of this classification number ("QA76.9.N38") represents the subject "Natural language processing". Thus, the Huang and Deng chapter, published in the second edition of *Handbook of Natural Language Processing*—as shown in the database record in Attachment I and the MARC bibliographic records in Attachments F, G, and H—was indexed according to its subject matter.

39. Based on my review of Exhibit 1014, my review of the bibliographic records in the online catalogs of the Library of Congress, the Indiana University system, and the Washington University in St. Louis, my review of the database record in the Web of Science, as well as my experience in library cataloging, it is my opinion that the Huang and Deng chapter, published in the second edition of *Handbook of Natural*

Language Processing, was publicly accessible at least as early as May 12, 2011, if not earlier than that. Accordingly, it is my opinion that the Huang and Deng chapter was accessible to the public prior to Nov. 17, 2011.

VII. "LECTURE 12: AN OVERVIEW OF SPEECH RECOGNITION" (EXHIBIT 1016)

40. I understand that a document titled "Lecture 12: An Overview of Speech Recognition" is to be submitted concurrently with this Declaration as Exhibit 1016.

41. Based on my review of Exhibit 1016, it appears on its face to be a class lecture titled "An Overview of Speech Recognition" (hereafter "Speech Recognition lecture"), and there is no author cited on the lecture. It appears to be from the University of Rochester Department of Computer Science website, given that the URL of "https://www.cs.rochester.edu" appears on the top of the first page of the exhibit. Exhibit 1016 was first captured by the Internet Archive's Wayback Machine on November 4, 2003 from the University of Rochester Department of Computer Science website. This is supported by the URL on the first page of the exhibit, which includes the URL of the University of Rochester Department of Computer Science website ("https://www.cs.rochester.edu/"). The first page of Exhibit 1016 is a screen capture of the first page of the lecture and information about the capture from the Internet Archive's Wayback Machine. Pages 2-9 of Exhibit 1016 include the lecture. I received a copy of this webpage from counsel, who, I understand, obtained Exhibit 1016 from the Internet Archive's Wayback Machine. I note that those parts of the text that comprise

Exhibit 1016 appear to be complete and there appear to be no visible alterations to the document.

42. The first page of Exhibit 1016 also appears to show that the lecture was captured by the Internet Archive's Wayback Machine on November 4, 2003. The is supported by the URL, which includes the November 4, 2003 date as part of the URL (20031104). To confirm this information, I reviewed the web-based archives of Exhibit 1016 on the Internet Archive's Wayback Machine, a widely-used database of web-based archives, as previously discussed in this declaration.

43. I entered the URL on top margin of first page of Exhibit 1016 in the search field of the Wayback Machine. I note that the lecture automatically opened when I searched the URL. Attached hereto as Attachment J is a true and correct copy of my search for Speech Recognition lecture on the Wayback Machine. It is the same lecture as Exhibit 1016, and includes the same title, URL, and content. This URL was first captured on November 4, 2003, and it has been captured by the Wayback Machine 24 times since then. Because the Wayback Machine captures publicly available webpages, it is my opinion that the Speech Recognition lecture was made publicly available to the public at least as early as November 4, 2003.

44. To support my opinion, I searched for the URL on the World Wide Web, and I found the Speech Recognition lecture at "https://www.cs.rochester.edu/u/james/CSC248/Lec12.pdf". I note that the lecture automatically opened when I searched for the URL. To find additional information

about the lecture, I cut text off the URL after each slash. First, I searched for "https://www.cs.rochester.edu/u/james/CSC248" and I found the Speech Recognition lecture on a website for a University of Rochester Department of Computer Science course numbered CSC 248/448. Attachment K is a true and correct copy of the "CSC 248/448 Home Page" website that I obtained from the World Wide Web. It includes a link to "Lecture 12: Approaches to Speech Recognition." I clicked on the URL, and the lecture is identical to the Speech Recognition lecture in Exhibit 1016. This website supports my opinion that the Speech Recognition lecture was available at least as early as November 4, 2003. The top of the "CSC 248/448 Home Page" website is dated "Fall 2003," indicating that the course was taught in the fall 2003 semester. At the bottom of the website, there are specific due dates for assignments. These dates all occur in fall 2003, and "Assignment 3" is due on November 4, 2003. This date matches the date the Wayback Machine captured the Speech Recognition lecture.

45. I also searched for the author of the Speech Recognition lecture by cutting more text off the URL. I searched for "https://www.cs.rochester.edu/u/james" and I found the webpage for James F. Allen, a Professor of Computer Science in the University of Rochester Department of Computer Science. Attachment L is a true and correct copy of the webpage for James F. Allen that I obtained from the World Wide Web. Because the URL captured by the Wayback Machine includes his name, it is reasonable to assume that he is responsible for teaching the CSC 248/448 course in fall

2003 at the University of Rochester Department of Computer Science, and that he is mostly likely the author of the Speech Recognition lecture.

46. Based on my review of Exhibit 1016, the web-based archives, the information available on the Web, as well as my experience in library cataloging, it is my opinion that the Speech Recognition lecture was available on the University of Rochester Department of Computer Science website at least as early as November 4, 2003. Accordingly, it is my opinion that the Speech Recognition lecture was accessible to the public prior to Nov. 17, 2011.

VIII. "USING BROAD PHONETIC GROUP EXERTS FOR IMPROVED SPEECH RECOGNITION" (EXHIBIT 1017)

47. I understand that a document titled "Using Broad Phonetic Group Experts for Improved Speech Recognition" is to be submitted concurrently with this Declaration as Exhibit 1017.

48. Based on my review, Exhibit 1017 is a journal article titled "Using Broad Phonetic Group Experts for Improved Speech Recognition" by Patricia Scanlon, Daniel P.W. Ellis, and Richard B. Reilly (hereafter "Scanlon, Ellis, and Reilly article"). The first page of Exhibit 1017 indicates that the article was published by IEEE in the journal *IEEE Transactions on Audio, Speech, and Language Processing* (ISSN 1558-7916), in the volume 15, number 3 (March 2007) issue, on pages 803-812. The first page of the exhibit indicates a copyright date of 2007. I received a copy of this journal article from counsel, who made the copy which comprises Exhibit 1017. I note that those parts of

the text that comprise Exhibit 1017 appear to be complete and there appear to be no visible alterations to the document.

49. *IEEE Transactions on Audio, Speech, and Language Processing* is considered a serial publication, specifically a journal or periodical. Each issue of *IEEE Transactions on Audio, Speech, and Language Processing* was mailed to libraries that subscribed to the print version of the journal and/or made available to libraries through a subscription to the electronic version of the journal (directly through IEEE or indirectly through a reference database). However, determining the public availability of journal issues in libraries can be difficult because libraries do not usually make this information public. However, the publisher's website often includes information about when journal issues, and the articles in them, were made available to the public.

50. Attachment M is a true and correct copy of the webpage for the Scanlon, Ellis, and Reilly article that I obtained from the IEEE Xplore database, which is an authoritative source for information about IEEE publications. The database record shows citation information that matches Exhibit 1017, including article title, authors, abstract, journal title, volume, number, and date. The database record also indicates that the "date of publication" of the article is February 20, 2007. This indicates that the Scanlon, Ellis, and Reilly article was made publicly available on February 20, 2007, the date that the Volume 15, number 3 (March 2007) issue of *IEEE Transactions on Audio, Speech, and Language Processing* was published.

51. My opinion is supported by the Declaration of Gordon MacPherson, a Director at IEEE, which was submitted separately as Exhibit 1039. Mr. MacPherson says that "Vol. 15, Issue 3 was published on February 20, 2007. Copies of this publication were made available on this date."

52. The IEEE Xplore database record and Mr. MacPherson both indicate that the Scanlon, Ellis, and Reilly article was published on February 20, 2007, so it is reasonable to assume that print subscribers, including libraries, would have been mailed the Volume 15, number 3 (March 2007) issue of *IEEE Transactions on Audio, Speech, and Language Processing* around that time. In my previous experience working as a professional cataloger, journal issues, such as from the *IEEE Transactions on Audio, Speech, and Language Processing*, are typically received at libraries approximately two months or so after the date of publication. It depends on when the publisher mailed the print issue to subscribers. Library staff members process such print issues (a process called "serials check-in") and then places them on the shelves of the library, where they are available to the public. This check-in process can take some time; I estimate that there could be a delay of about one to three weeks (or more) after a print issue is first received in the library until it is "shelved" and available to the public. However, libraries that subscribe to the electronic version of the journal through IEEE would have had access to the Scanlon, Ellis, and Reilly article on February 20, 2007.

53. To further substantiate my opinion, I looked on the online public catalog of the United States Copyright Office. Attachment N is a true and correct copy of the

bibliographic record for *IEEE Transactions on Audio, Speech, and Language Processing* that shows cataloging information for the journal, as well as issues that have been registered by the United States Copyright Office. The bibliographic identifiers for the journal match those of the journal information in Exhibit 1017, including journal title, ISSN, and publisher. It also includes information about the volume 15, number 3 (March 2007) issue, in which the Scanlon, Ellis, and Reilly article was published. There are two entries, presumably one for the print version and one for online version of the issue. The record shows that the volume 15, number 3 (March 2007) issue was published on February 14, 2007. This is 6 days before the publisher indicates the issue was published and made publicly available. In addition, the copyright of the issue was registered on June 12, 2007 and June 22, 2007. This is around 4 months after the publisher indicates the issue was published and made publicly available.

54. Furthermore, the database record in Attachment M includes 27 indexing descriptors that provide subject access to the Scanlon, Ellis, and Reilly article. There are 9 "IEEE Keywords": Speech recognition, Mutual information, Context modeling, Feature extraction, Data mining, Time frequency analysis, Acoustic signal detection, Detectors, Error analysis. There are 4 "INSPEC: Controlled Indexing" descriptors: error statistics, signal classification, speech recognition, time-frequency analysis. There are 10 "INSPEC: Non-Controlled Indexing" descriptors: broad phonetic group experts, speech recognition, phoneme recognition, acoustic models, temporal context window, time-frequency feature set, phone recognition system, baseline classifier, TIMIT

continuous speech corpus, error rate reductions. There are 4 "Author Keywords": Automatic speech recognition, broad phonetic groups (BPGs), mixture of experts, mutual information (MI). Thus, the Scanlon, Ellis, and Reilly article—as shown in the database record in Attachment M—was indexed according to its subject matter.

55. Accordingly, based on my review of Exhibit 1017, the database record on the IEEE Xplore database, the Declaration of Gordon MacPherson, the cataloging information from the United States Copyright Office, as well as my experience in library cataloging, it is my opinion that the volume 15, number 3 (March 2007) issue of *IEEE Transactions on Audio, Speech, and Language Processing*, which includes the Scanlon, Ellis, and Reilly article, was published and made available on February 20, 2007. I estimate that it would have been made publicly available in libraries at least as early as March 20, 2007, if not earlier. Accordingly, it is my opinion that the Scanlon, Ellis, and Reilly article was accessible to the public prior to Nov. 17, 2011.

IX. "SPEECH PERCEPTION USING REAL-TIME PHONEME DETECTION: THE BEBE SYSTEM" (EXHIBIT 1018)

56. I understand that a document titled "Speech Perception Using Real-Time Phoneme Detection: The BeBe System" is to be submitted concurrently with this Declaration as Exhibit 1018.

57. Based on my review of Exhibit 1018, it appears on its face to be a technical report titled "Speech Perception Using Real-Time Phoneme Detection: The BeBe System" written by Latanya Sweeney and Patrick Thompson, with a copyright date of

1997 (hereafter "Sweeney and Thompson report"), as part of their work for the Laboratory for Computer Science at the Massachusetts Institute of Technology (MIT). I received a copy of this technical report from counsel, who, I understand, obtained Exhibit 1018 from the Internet Archive's Wayback Machine. I note that those parts of the text that comprise Exhibit 1018 appear to be complete and there appear to be no visible alterations to the document.

58. According to Latanya Sweeney's online Curriculum Vitae (https://latanyasweeney.org/cv.html), and shown in the screen capture below, the Sweeney and Thompson report is considered "Tech Report MIT-LCS-TR-736" from the "Massachusetts Institute of Technology, Laboratory for Computer Science." It also includes a URL to the technical report on a website called the Data Privacy Lab.

101. *Speech perception using real-time phoneme detection, the BeBe System.*
Sweeney, L. and Thompson, P.
Massachusetts Institute of Technology, Laboratory for Computer Science: Tech Report MIT-LCS-TR-736. 1998.
dataprivacylab.org/projects/bebe/index.html

59. Attachment O is a true and correct copy of a webpage from the Data Privacy Lab that includes citation information and an abstract for the Sweeney and Thompson report. The title, authors, and abstract match Exhibit 1018, and the citation information confirms that the Sweeney and Thompson report is considered "Tech Report MIT-LCS-TR-736" from the "Massachusetts Institute of Technology, Laboratory for Computer Science." The webpage shows a 1998 date for the report, but the PDF file available on the webpage is the same report as in Exhibit 1018, with a

copyright date of 1997. Attachment P is a true and correct copy of the Sweeney and Thompson report available on the Data Privacy Lab webpage.

60. The Data Privacy Lab webpage also shows a copyright date of 2011 at the bottom of the webpage. This indicates the year the Sweeney and Thompson report was made available through the Data Privacy Lab, but with a copyright date of 1997, the Sweeney and Thompson report was most likely made publicly available before 2011.

61. I searched for the URL "publications.csail.mit.edu/lcs/pubs/pdf/MIT-LCS-TR-736.pdf" on the Internet Archive's Wayback Machine, a widely-used database of web-based archives, as previously discussed in this declaration. Attachment Q is a true and correct copy of the results of this search. It shows that the URL "http://publications.csail.mit.edu/lcs/pubs/pdf/MIT-LCS-TR-736.pdf" was captured three times between July 12, 2010 and January 21, 2022. I note that the Sweeney and Thompson report automatically opened on my computer after I searched for the URL, and I verified that the Sweeney and Thompson report in the file is the same as in Exhibit 1018. As shown in Attachment Q, the Wayback Machine captured the Sweeney and Thompson report on July 12, 2010, which means the report was publicly available at least by that time.

62. To support my opinion, I searched for the URL on the World Wide Web, and I found the same Sweeney and Thompson report. To find additional information about the report, I cut text off the URL after each slash. First, I searched for "publications.csail.mit.edu/lcs/pubs/pdf/" and found a "Index of lcs/pubs/pdf" with a list

of many files that are included at that URL. Attachment R is a true and correct copy of screen captures taken from this index. As shown in the first screen capture in Attachment R, the index includes each file name and the date each file was last modified. I scrolled down to "MIT-LCS-TR-736.pdf", as shown in the second screen capture of Attachment R, and found that the PDF file that contains the Sweeney and Thompson report was last modified on February 16, 2001. This means that the PDF file at the URL "http://publications.csail.mit.edu/lcs/pubs/pdf/MIT-LCS-TR-736.pdf" has not been changed since February 16, 2001. The Sweeney and Thompson report has most likely been available at that URL since then, and it is still publicly available, as evidenced by the January 21, 2022 capture taken by the Wayback Machine, as detailed at the top of Attachment Q.

63. My opinion is further supported by the Affidavit of Nathaniel E. Frank-White, a Records Request Processor at the Internet Archive, which was submitted separately as Exhibit 1041. In his Attachment B3, Mr. Frank-White confirms that Exhibit 1018 was first captured by the Internet Archive's Wayback Machine. The title, authors, date, content in Mr. Frank-White's Appendix B3 match the title, authors, date, and content in Exhibit 1018 and this Declaration's Attachment P.

64. Based on my review of Exhibit 1018, the web-based archives, the information available on the Web, the Affidavit of Nathaniel E. Frank-White, as well as my experience in library cataloging, it is my opinion that the Sweeney and Thompson report was publicly available on the Web at least as early as July 12, 2010, and most

likely much earlier than that. Accordingly, it is my opinion that the Sweeney and Thompson report was accessible to the public prior to Nov. 17, 2011.

X. "VOICE-TO-PHONEME CONVERSION ALGORITHMS FOR VOICE-TAG APPLICATIONS IN EMBEDDED PLATFORMS" (EXHIBIT 1019)

65. I understand that a document titled "Voice-to-Phoneme Conversion Algorithms for Voice-Tag Applications in Embedded Platforms" is to be submitted concurrently with this Declaration as Exhibit 1019.

66. Based on my review, Exhibit 1019 is a journal article titled "Voice-to-Phoneme Conversion Algorithms for Voice-Tag Applications in Embedded Platforms" by Yan Ming Cheng, Changxue Ma, and Lynette Meinar (hereafter "Cheng, Ma, and Meinar article"). The first page of Exhibit 1019 indicates the article was published by Hindawi Publishing Corporation in the journal *EURASIP Journal on Audio, Speech, and Music Processing*, in "Volume 2008" and has an Article ID number of "568737". It has a copyright date of 2008. I received a copy of this journal article from counsel, who made the copy which comprises Exhibit 1019. I note that those parts of the text that comprise Exhibit 1019 appear to be complete and there appear to be no visible alterations to the document.

67. *EURASIP Journal on Audio, Speech, and Music Processing* is considered a serial publication, specifically a journal or periodical. It is an open-access journal, which means that its articles are publicly available online on the day they are published; there is no corresponding print version of the journal. The first page of the exhibit confirms

that the Cheng, Ma, and Meinar article in Exhibit 1019 is an open-access article: "This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited." I confirmed that *EURASIP Journal on Audio, Speech, and Music Processing* published open-access articles by looking on the DOAJ: Directory of Open Access Journals, an authoritative database that catalogs and indexes open-access journals. Attachment S is a true and correct copy of the webpage for *EURASIP Journal on Audio, Speech, and Music Processing* taken from the DOAJ: Directory of Open Access Journals. It says the journal "began publishing in open access in 2006."

68. Libraries provide access to open access journals by linking directly to the journal's website or through reference databases, and a library may provide access to a journal title through multiple databases, depending upon how many databases a library has licensed. For example, Attachment T is a true and correct copy of the cataloging information about *EURASIP Journal on Audio, Speech, and Music Processing* from the University of Alabama online library catalog, which shows that the library provides access to the journal through three different databases: Springer Nature Compete Journals, DOAJ: Directory of Open Access Journals, and EBSCO Open Access Journals.

69. Because *EURASIP Journal on Audio, Speech, and Music Processing* is an open-access journal with articles available only online, it is not possible to determine

the public availability of the Cheng, Ma, and Meinar article through library records. Instead, this information is provided on the article's webpage. Attached hereto as Attachment U is a true and correct copy of the webpage for the Cheng, Ma, and Meinar article available through SpringerOpen website. SpringerOpen is latest publisher of the journal. As shown in Attachment U, the title, authors, article number, abstract, text, and so on match the information in Exhibit 1019. "Published: 04 October 2007" is at the top of the webpage, which indicates that the Cheng, Ma, and Meinar article was published and made publicly available online on October 4, 2007.

70. Furthermore, the last page of the database record in Attachment U includes 5 indexing descriptors that provide subject access to the Cheng, Ma, and Meinar article: Speech Recognition, Target Language, Dynamic Time Warping, Acoustic Model, Source Language. In addition, because this is an open-access journal article, users can search the full-text of this article using any search terms. Thus, the Cheng, Ma, and Meinar article—as shown in the database record in Attachment U—was indexed according to its subject matter.

71. Accordingly, based on my review of Exhibit 1019, the information from DOAJ: Directory of Open Access Journals, the University of Alabama online library catalog, the SpringerOpen website, as well as my experience in library cataloging, it is my opinion that the Cheng, Ma, and Meinar article was available to the public at least as early as October 4, 2007. Accordingly, it is my opinion that the Cheng, Ma, and Meinar article was accessible to the public prior to Nov. 17, 2011.

XI. "NETWORK+ STUDY GUIDE, DELUXE EDITION" (EXHIBIT 1038)

72. I understand that a document titled "Network+ Study Guide, Deluxe Edition" is to be submitted concurrently with this Declaration as Exhibit 1038.

73. Based on my review, Exhibit 1038 appears to be parts of a book titled "Network+ Study Guide, Deluxe Edition" written by David Groth, Toby Skandier, Todd Lammle, and William Tedder, published by Sybex (hereafter "Groth et al. book"). Exhibit 1038 is a copy of parts of the Groth et al. book made from a print copy obtained from the Library of Congress, including: the front cover and spine, back cover, two pages of marketing materials, a copy of "CD 1 – Study Tools", four pages of "Objectives", half title page, title page, title page verso, letter from the publisher, software license agreement, authors' quotes, acknowledgments, contents at a glance, full contents on pages [x]-xvii, page xxxviii of the introduction, chapter 1 on pages [1]-43, and the first page of chapter 2. The book's title page verso shows a copyright date of 2005; a 10-digit ISBN: 0-7821-4416-0; a Library of Congress Card Number (today, called a Library of Congress Control Number): 2005920773; and bears a stamp indicating Library of Congress Copyright Office July 13, 2005. The spine bears a call number label: TK5105.5 .N46673 2005 Copy 1. The back cover bears a barcode label: Library of Congress 0 011 405 276 0. I received a copy Exhibit 1038 from counsel, who, I understand, obtained the copy from the Library of Congress. I note that those parts of the text that comprise Exhibit 1038 appear to be complete and there appear to be no visible alterations to the document.

74. Attached hereto as Attachment V is a true and correct copy of the online catalog record for the Groth et al. book from the Library of Congress online catalog that I obtained. The document cataloged in this record is the Groth et al. book as verified by the fields listing the title: Network+ study guide; edition: Deluxe ed.; publisher and copyright date: Sybex in 2005; 10-digit ISBN: 0782144160; and, LCCN: 2005920773. I also compared the LC Classification/Call Number (TK5105.5 .N46673 2005) with that shown on the spine of Exhibit 1038, and the values are equivalent on both. The record indicates that two print copies of the book are held by the Library of Congress. The Attachment indicates that the Groth et al. book as cataloged at the Library of Congress is currently held by that institution. All identifying characteristics, including stamps and notations, on Exhibit 1038 are consistent with library practice and procedure that I have observed during my career as a professional librarian. Based on my review of Exhibit 1038, in my opinion, Exhibit 1038 is a true and correct copy in a condition that creates no suspicion about its authenticity.

75. Attached hereto as Attachment W is a true and correct copy of the MARC bibliographic record for the Groth et al. book that I obtained from the online library catalog of the Library of Congress. The bibliographic identifiers match those of Exhibit 1038, including title (field 245), edition (field 250), publisher and date (field 260), 10-digit ISBN (field 020), and Library of Congress Card Number (field 010). Field 008 on the bibliographic record shows that the bibliographic record was created on January 18, 2005. Field 005 explains when the bibliographic record in the Library of Congress's

online catalog was last updated, which was June 25, 2007. The 955 field on this bibliographic show the dates when the book moved from different people/departments at the Library of Congress. The latest date is June 25, 2007, and the information indicates when a second copy of the book went to the Library of Congress's Binding and Collections Care Division ("2007-06-25 Copy 2 to BCCD"). Because copy 1 of the Groth et al. book at the Library of Congress is date stamped July 13, 2005 (Exhibit 1038), and the MARC bibliographic record shows that a second copy of the Groth et al. book was at the Library of Congress on June 25, 2007, one can reasonably assume that the book was made publicly available between July 13, 2005 and June 25, 2007. It is my opinion that the Groth et al. book was published and accessible to the public at least as early as June 25, 2007.

76. My opinion is further substantiated by bibliographic records in other libraries that hold the Groth et al. book. Attached hereto as Attachment X is a true and correct copy of the MARC bibliographic record for the Groth et al. book from the online library catalog of Alabama State University that I obtained. The bibliographic identifiers match those of Exhibit 1038, including title (field 245), edition (field 250), publisher and date (field 260), 10-digit ISBN (field 020), and Library of Congress Card Number (field 010). Field 008 on the bibliographic record shows that the bibliographic record was created on January 18, 2005, downloaded from the OCLC bibliographic database, as indicated in field 035. Field 005 explains when the bibliographic record in the Alabama State University online library catalog was last updated, which was May

11, 2007. This date is an indication when the Groth et al. book was cataloged at the Alabama State University Library, although it may have taken several months for the item to be barcoded and shelved. This is around 1 month before the second copy of the Groth et al. book was at the Library of Congress.

77. Attached hereto as Attachment Y is a true and correct copy of the MARC bibliographic record for the Groth et al. book from the online library catalog of the British Library that I obtained. The bibliographic identifiers match, or are equivalent to, those of Exhibit 1038, including title (field 245), edition (field 250), publisher and date (field 260), 10-digit ISBN (field 020), and Library of Congress Card Number (field 010). Field 008 on the bibliographic record shows that the bibliographic record was created on November 17, 2004. Field 005 explains when the bibliographic record in the British Library online catalog was last updated, which was November 2, 2005. This date is an indication when the Groth et al. book was cataloged at the British Library, although it may have taken several months for the item to be barcoded and shelved. This is around one year and 7 months before the second copy of the Groth et al. book was at the Library of Congress.

78. Furthermore, the bibliographic records in Attachments W and X include three subject headings in field 650 reading "Computer networks |x Examinations |v Study guides", "Telecommunications engineers |x Certification |v Study guides", "Microsoft software |x Examinations |v Study guides". The “_0” in field 650 indicates the subject heading was taken from the *Library of Congress Subject Headings*. In

addition, Attachments W and X include a *Library of Congress Classification* number based on subject in field 050 (“TK5105.5 .N46673 2005” and "TK5105.5 .N47 2005") The first part of these classification numbers (“TK5105.5”) represents the subject “Computer networks" and is assigned to general works about the subject. Attachment Y includes similar subject access to Attachments W and X. The record includes two subject headings from the *Library of Congress Subject Headings*, "Computer networks |x Examinations |v Study guides" and "Electronic data processing personnel |x Certification |v Study guides". There is also a *Dewey Decimal Classification* number assigned in field 082 ("004.6"). This represents the subject "Computer science - Interfacing and communications". Thus, the Groth et al. book—as shown in the MARC bibliographic records in Attachments W, X, and Y—was indexed according to its subject matter.

79. Based on my review of Exhibit 1038, my review of the bibliographic record in the online library catalog of the Library of Congress, the bibliographic record in the online library catalog of Alabama State University, the bibliographic record in the online library catalog of the British Library, as well as my experience in library cataloging, it is my opinion that the Groth et al. book was publicly available at least as early as June 25, 2007, and most likely much earlier than that. Accordingly, it is my opinion that the Groth et al. book was accessible to the public prior to Nov. 17, 2011.

XII. CONCLUSION

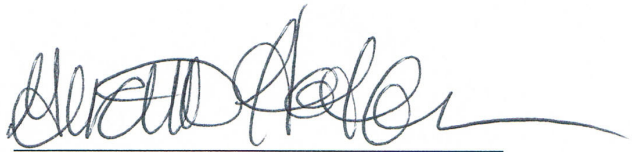
80. In signing this declaration, I recognize that the declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I also recognize that I may be subject to cross-examination in the case and that cross-examination will take place within the United States. If cross-examination is required of me, I will appear for cross-examination within the United States during the time allotted for cross-examination.

81. If asked to more fully explain my opinions as expressed in this declaration, I reserve the right to rely on various portions of the references and library catalogs that I have discussed in this report, but have not explicitly cited.

82. I reserve the right to supplement my opinions in the future to respond to any arguments that the Patent Owner raises and to address additional information as it becomes available to me.

83. I hereby declare that all statements made herein of my own knowledge are true, and that all statements made on information and belief are believed to be true and that such statements are made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Date: July 18, 2022



Dr. Gretchen Hoffman, Ph.D.

ATTACHMENT

A



ScienceDirect

Encyclopedia of Language & Linguistics (Second Edition)
2006, Pages 806-819

Speech Recognition, Automatic: History

B.-H. Juang, L.R. Rabiner

Show more

Outline | Share Cite

https://doi.org/10.1016/B0-08-044854-2/00906-8 Get rights and content

Designing a machine that mimics human behavior, particularly the capability of speaking naturally and responding properly to spoken language, has intrigued engineers and scientists for centuries. Since the 1930s, when Homer Dudley of Bell Laboratories proposed a system model for speech analysis and synthesis (Dudley, 1939; Dudley *et al.*, 1939), the problem of automatic speech recognition has been approached progressively, from a simple machine that responds to a small set of sounds to a sophisticated system that responds to fluently spoken natural language. Based on major advances in statistical modeling of speech in the 1980s, automatic speech recognition systems today find widespread application in tasks that require a human-machine interface. Examples are automatic call processing in the telephone network and query-based information systems that provide updated travel information, stock price quotations, weather reports, etc. In this article we review some major highlights in the research and development of automatic speech recognition during the last few decades to provide a technological perspective and an appreciation of the fundamental progress that has been made in this important area of information and communication technology.

acoustic modeling; automatic transcription; dialog systems; finite state network; hidden Markov models; keyword spotting; language modeling; neural networks; office automation; pattern recognition; spectral analysis; Speech recognition; speech

IPR2022-01289 Page 00043

understanding; statistical modeling; time normalization

Recommended articles

Cited by (0)



Biing-Hwang Juang (Ph.D., University of California, Santa Barbara, 1981) has worked at Speech Communications Research Laboratory and Signal Technology, Inc. on several government-sponsored research projects. He joined the Acoustics Research Department at Bell Laboratories in 1982. In 1996 he became the director of the Acoustics and Speech Research Department at Bell Labs and, in 2001, director of Multimedia Technologies at Avaya Labs (a spin-off of Bell Labs). He joined the Georgia Institute of Technology faculty in 2002 as Motorola Foundation Chair Professor in the School of Electrical and Computer Engineering. His publications include *Fundamentals of speech recognition* (with L R Rabiner, 1993). He holds about 20 patents, has served as editor-in-chief of *IEEE Transactions on Speech and Audio Processing,* and is the recipient of numerous professional awards.



Lawrence R Rabiner received his Ph.D. in electrical engineering from MIT in 1967. From 1962 to 1964 he participated in the cooperative program in electrical engineering at AT&T Bell Labs, Whippany and Murray Hill, New Jersey, working on digital circuitry, military communications problems, and binaural hearing. He held various positions at AT&T Bell Labs and AT&T Labs, culminating in vice president of research from 1998 through 2002, when he retired. He is now professor of electrical and computer engineering at Rutgers University and the associate director of the Center for Advanced Information Processing there. He holds a joint appointment as professor of electrical and computer engineering at the University of California at Santa Barbara. A leading specialist in digital speech processing, he is coauthor of *Theory and application of digital signal processing* (Prentice-Hall, 1975), *Digital processing of speech signals* (Prentice-Hall, 1978), *Multirate digital signal processing* (Prentice-Hall, 1983), and *Fundamentals of speech recognition* (Prentice-Hall, 1993).

IPR2022-01289 Page 00044

Copyright © 2006 Elsevier Ltd. All rights reserved.



Copyright © 2022 Elsevier B.V. or its licensors or contributors.
ScienceDirect® is a registered trademark of Elsevier B.V.

RELX™

ATTACHMENT

B

MARC bibliographic record for *Encyclopedia of Language & Linguistics*, including the Juang and Rabiner article, in the shared online library catalog of the Indiana University system. Available: https://iucat.iu.edu/catalog/6559065

Librarian View

LEADER 00000cam a2200000 a 45 0	
001 6559065	
003 OCoLC	
005 20070303155644.9	
008 050428s2006 mauabc be 001 0 eng||	
035	a| (Sirsi)6559065
035	a| (OCoLC)ocm61441874
040	a| UKM b| eng c| UKM d| BAKER d| BOS d| CBC d| DAY d| NLGGC d| MUQ d| YDXCP
015	a| GBA572244 2| bnb
016 7	a| 013280004 2| Uk
020	a| 0080442994 (set)
020	a| 9780080442990 (set)
020	a| 0080443575 (v. 1)
020	a| 9780080443577 (v. 1)
020	a| 0080443583 (v. 2)
020	a| 9780080443584 (v. 2)
020	a| 0080443591 (v. 3)
020	a| 9780080443591 (v. 3)
020	a| 0080443605 (v. 4)
020	a| 9780080443607 (v. 4)
020	a| 0080443613 (v. 5)
020	a| 9780080443614 (v. 5)
020	a| 0080443621 (v. 6)
020	a| 9780080443621 (v. 6)
020	a| 008044363X (v. 7)
020	a| 9780080443638 (v. 7)
020	a| 0080443648 (v. 8)
020	a| 9780080443645 (v. 8)
020	a| 0080443656 (v. 9)
020	a| 9780080443652 (v. 9)
020	a| 0080443664 (v. 10)
020	a| 9780080443669 (v. 10)

020			a| 0080443672 (v. 11)
020			a| 9780080443676 (v. 11)
020			a| 0080443680 (v. 12)
020			a| 9780080443683 (v. 12)
020			a| 0080443699 (v. 13)
020			a| 9780080443690 (v. 13)
020			a| 0080443702 (v. 14)
020			a| 9780080443706 (v. 14)
029	1		a| NLGGC b| 267764626
029	1		a| YDXCP b| 2163783
042			a| pcc
082	0	4	a| 403 2| 22
084			a| 17.01 2| bcl
090			a| P29 b| .E48 2006
049			a| I3UA
245	0	0	a| Encyclopedia of language & linguistics / c| Keith Brown, editor-in-chief ; co-ordinating editors, Anne H. Anderson ... [et al.].
246	3		a| Encyclopedia of language and linguistics
250			a| 2nd ed.
260			a| Boston : b| Elsevier, c| c2006.
300			a| 14 v. : b| ill., maps (some col.), ports. ; c| 29 cm.
500			a| Previous ed.: Oxford ; New York : Pergamon, 1994.
504			a| Includes bibliographical references and index.
505	0		a| v. 1. A-Bil -- v. 2. Bil-Con -- v. 3. Con-Dt -- v. 4. Du-Gen -- v. 5. Gen-Int -- v. 6. Inu-Lea -- v. 7. Leb-Mei -- v. 8. Mel-N -- v. 9. O-Pou -- v. 10. Pou-Sca -- v. 11. Sca-Spe -- v. 12. Spe-Top -- v. 13. Tor-Z -- v. 14. Appendices, Index.
650		0	a| Language and languages v| Encyclopedias. =| ^A1650517
650		0	a| Linguistics v| Encyclopedias. =| ^A1650518
650	1	7	a| Taalwetenschap. 2| gtt
650	1	7	a| Taal. 2| gtt
650		6	a| Langage et langues v| Encyclopédies.
650		6	a| Linguistique v| Encyclopédies.
700	1		a| Brown, E. K. =| ^A161300
700	1		a| Anderson, Anne, d| 1954- =| ^A903395
596			a| 1 46

999	a| P29 .E48 2006 v.1 i| 30000109354559 j| P 000029 .E48 2006 V.000001 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC z| CAT;08-23-2006;CC333
999	a| P29 .E48 2006 v.2 i| 30000109354567 j| P 000029 .E48 2006 V.000002 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.3 i| 30000109354575 j| P 000029 .E48 2006 V.000003 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.4 i| 30000109354583 j| P 000029 .E48 2006 V.000004 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.5 i| 30000109354591 j| P 000029 .E48 2006 V.000005 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.6 i| 30000109354500 j| P 000029 .E48 2006 V.000006 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.7 i| 30000109354518 j| P 000029 .E48 2006 V.000007 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.8 i| 30000109354526 j| P 000029 .E48 2006 V.000008 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.9 i| 30000109354534 j| P 000029 .E48 2006 V.000009 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.10 i| 30000109354542 j| P 000029 .E48 2006 V.000010 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.11 i| 30000109354468 j| P 000029 .E48 2006 V.000011 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.12 i| 30000109354476 j| P 000029 .E48 2006 V.000012 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.13 i| 30000109354484 j| P 000029 .E48 2006 V.000013 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC
999	a| P29 .E48 2006 v.14 i| 30000109354492 j| P 000029 .E48 2006 V.000014 l| _RCREFDPT m| B-WELLS n| _RCREFDPT w| LC

999	a| P29 .E48 2006 v.1 i| 30000110755927 j| P 000029 .E48 2006 V.000001 l| REFERENCE m| SBEND n| REFERENCE w| LC z| CAT;06-20-2007;SB159
999	a| P29 .E48 2006 v.2 i| 30000110770470 j| P 000029 .E48 2006 V.000002 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.3 i| 30000110770488 j| P 000029 .E48 2006 V.000003 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.4 i| 30000110770496 j| P 000029 .E48 2006 V.000004 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.5 i| 30000110770504 j| P 000029 .E48 2006 V.000005 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.6 i| 30000110770512 j| P 000029 .E48 2006 V.000006 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.7 i| 30000110771361 j| P 000029 .E48 2006 V.000007 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.8 i| 30000110771379 j| P 000029 .E48 2006 V.000008 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.9 i| 30000110771387 j| P 000029 .E48 2006 V.000009 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.10 i| 30000110771395 j| P 000029 .E48 2006 V.000010 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.11 i| 30000110771403 j| P 000029 .E48 2006 V.000011 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.12 i| 30000110771411 j| P 000029 .E48 2006 V.000012 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.13 i| 30000110771429 j| P 000029 .E48 2006 V.000013 l| REFERENCE m| SBEND n| REFERENCE w| LC
999	a| P29 .E48 2006 v.14 i| 30000110771437 j| P 000029 .E48 2006 V.000014 l| REFERENCE m| SBEND n| REFERENCE w| LC

ATTACHMENT

C

UNIVERSITY OF MICHIGAN LIBRARY EXPLORE

SEARCH

Account My Favorites Log in ANN ARBOR FLINT

Keyword

encyclopedia of language and linguistics

Advanced

Search Tip: Enter one or more keywords. Use quotes to search for a phrase (e.g. solar power; polar bears; "systems of oppression"). See tips about Basic Keyword Searching.

Everything Catalog Articles Databases Online Journals Guides and More

Catalog / Record

Book Encyclopedias

The encyclopedia of language and linguistics / Keith Brown, editor-in-chief.

Record info:

Format
Book
Encyclopedias

Related Title
Asher, R. E. Encyclopedia of language and linguistics.
Simpson, J. M. Y. Encyclopedia of language and linguistics.

Other Titles
LAnguage and linguistics

Contributors
Brown, E. K.
Asher, R. E.
Simpson, J. M. Y.
ScienceDirect (Online service)

Published/Created
Amsterdam ; Oxford : Elsevier, 2006

Edition
2nd ed.

Ask a Librarian

Language
English

Copy-Specific Note
A gift from Dr. Carl Winberg in memory of his mother, Marian Davis Winberg.

Note
Previous ed.: published as by R.E. Asher and J.M.Y. Simpson. New York; Oxford: Pergamon, 1994.
Includes index.

Physical Description
14 v. : ill., maps (some col.), ports. ; 29 cm.

Media Format
Mode of access: Internet.

ISBN
0080442994
0080443575
0080443583
0080443591

Show all 15 ISBNs

OCLC Number
61441874

Subjects (LCSH)
Language and languages -- Encyclopedias.
Linguistics -- Encyclopedias.

Contents Listing
v. 1. A-Bil -- v. 2. Bil-Con -- v. 3. Con-Dt -- v. 4. Du-Gen -- v. 5. Gen-Int -- v. 6. Inu-Lea -- v. 7. Leb-Mei -- v. 8. Mel-N -- v. 9. O-Pou -- v. 10. Pou-Sca -- v. 11. Sca-Spe -- v. 12. Spe-Top -- v. 13. Tor-Z -- v. 14. Appendices, Index.

The University of Michigan Library aims to describe library materials in a way that respects the people and communities who create, use, and are represented in our collections. Report harmful or offensive language in catalog records, finding aids, or elsewhere in our collections anonymously through our metadata feedback form. More information at Remediation of Harmful Language.

Available at:

Online Resources · 1 item

HathiTrust Digital Library · 14 items

Hatcher Graduate Reference Rm. - 2nd floor N · 14 items

Actions Select what to do with this record.

Ask a Librarian

IPR2022-01289 Page 00053

Email Text Citation Export Citation (EndNote) Export Citation (RIS) Copy link

MARC Data

LDR		00000nam a2200325 a 4500
001		990050989980106381
005		20201012115623.0
006		m d
007		cr bn ---auaua
008		050428s2006 ne e 001 0 eng
035		**|a** (MiU)005098998MIU01
015		**|a** GBA572244 **|2** bnb
016	7	**|a** 013280004 **|2** Uk
020		**|a** 0080442994 **|q** (set) **|0** (uri) http://www.isbnsearch.org/isbn/0080442994
020		**|a** 0080443575 **|q** (v. 1) **|0** (uri) http://www.isbnsearch.org/isbn/0080443575
020		**|a** 0080443583 **|q** (v. 2) **|0** (uri) http://www.isbnsearch.org/isbn/0080443583
020		**|a** 0080443591 **|q** (v. 3) **|0** (uri) http://www.isbnsearch.org/isbn/0080443591
020		**|a** 0080443605 **|q** (v. 4) **|0** (uri) http://www.isbnsearch.org/isbn/0080443605
020		**|a** 0080443613 **|q** (v. 5) **|0** (uri) http://www.isbnsearch.org/isbn/0080443613
020		**|a** 0080443621 **|q** (v. 6) **|0** (uri) http://www.isbnsearch.org/isbn/0080443621
020		**|a** 008044363X **|q** (v. 7) **|0** (uri) http://www.isbnsearch.org/isbn/008044363X
020		**|a** 0080443648 **|q** (v. 8) **|0** (uri) http://www.isbnsearch.org/isbn/0080443648
020		**|a** 0080443656 **|q** (v. 9) **|0** (uri) http://www.isbnsearch.org/isbn/0080443656
020		**|a** 0080443664 **|q** (v. 10) **|0** (uri) http://www.isbnsearch.org/isbn/0080443664
020		**|a** 0080443672 **|q** (v. 11) **|0** (uri) http://www.isbnsea

Ask a Librarian

020			**|a** 0080443680 **|q** (v. 12) **|0** (uri) http://www.isbnsearch.org/isbn/0080443680
020			**|a** 0080443699 **|q** (v. 13) **|0** (uri) http://www.isbnsearch.org/isbn/0080443699
020			**|a** 0080443702 **|q** (v. 14) **|0** (uri) http://www.isbnsearch.org/isbn/0080443702
035			**|a** (OCoLC)ocm61441874 **|0** (uri) http://www.worldcat.org/oclc/61441874
040			**|a** UKM **|c** UKM **|d** BAKER
049			**|a** EYM+
050	0	0	**|a** P29 **|b** .E48 2006
245	0	4	**|a** The encyclopedia of language and linguistics / **|c** Keith Brown, editor-in-chief.
246	3	0	**|a** LAnguage and linguistics
250			**|a** 2nd ed.
260			**|a** Amsterdam ; **|a** Oxford : **|b** Elsevier, **|c** 2006.
300			**|a** 14 v. : **|b** ill., maps (some col.), ports. ; **|c** 29 cm.
500			**|a** Previous ed.: published as by R.E. Asher and J.M.Y. Simpson. New York; Oxford: Pergamon, 1994.
500			**|a** Includes index.
505	0		**|a** v. 1. A-Bil -- v. 2. Bil-Con -- v. 3. Con-Dt -- v. 4. Du-Gen -- v. 5. Gen-Int -- v. 6. Inu-Lea -- v. 7. Leb-Mei -- v. 8. Mel-N -- v. 9. O-Pou -- v. 10. Pou-Sca -- v. 11. Sca-Spe -- v. 12. Spe-Top -- v. 13. Tor-Z -- v. 14. Appendices, Index.
590			**|a** A gift from Dr. Carl Winberg in memory of his mother, Marian Davis Winberg.
538			**|a** Mode of access: Internet.
650		0	**|a** Language and languages **|v** Encyclopedias. **|0** (uri) http://id.loc.gov/authorities/subjects/sh85074518
650		0	**|a** Linguistics **|v** Encyclopedias. **|0** (uri) http://id.loc.gov/authorities/subjects/sh85077222
700	1		**|a** Brown, E. K. **|0** (uri) http://id.loc.gov/authorities/names/n80123959 **|0** (uri) http://viaf.org/viaf/sourceID/LC|n80123959
700	1		**|a** Asher, R. E. **|t** Encyclopedia of language and linguistics.
700	1		**|a** Simpson, J. M. Y. **|t** Encyclopedia of language and

Ask a Librarian

710	2		**|a** ScienceDirect (Online service) **|0** (uri) http://id.loc.gov/authorities/names/no00037061 **|0** (uri) http://viaf.org/viaf/sourceID/LC|no00037061
856	4	1	**|z** Access to the ScienceDirect online version restricted; authentication may be required: **|u** http://www.sciencedirect.com/science/referenceworks/9780080448541 **|m** ScienceDirect
997			**|a** MARS
998			**|c** css ya ImmCat 2006-02-07
998			**|c** wheigel-rdu LSPMigrProj 20201012 ELEC
958			**|a** MiU
995			**|a** 20
BIB			**|u** 2021-07-18 04:23:18 US/Eastern **|c** 2021-06-21 19:47:21 US/Eastern **|s** false
E56			**|m** ScienceDirect **|1** ELEC **|t** static **|u** https://na04.alma.exlibrisgroup.com/view/uresolver/01UMICH_INST /openurl?u.ignore_date_coverage=true&portfolio_pid=531030323560006381& Force_direct=true **|a** 2021-06-23 22:13:47 **|v** https://umich.alma.exlibrisgroup.com/view/uresolver/01UMICH_INST /openurl?&rfr_id=info:sid/primo.exlibrisgroup.com& svc_dat=CTO?u.ignore_date_coverage=true?u.ignore_date_coverage=true& rft.mms_id=990050989980106381 **|z** Access to the ScienceDirect online version restricted; authentication may be required: **|q** true **|s** Available **|8** 531030323560006381
852	0		**|b** HATCH **|a** MiU **|c** GLRF **|h** P 29 .E481 2006 **|8** 22998095460006381
866	4	0	**|a** (v. 1-14 2006) **|8** 22998095460006381
974			**|8** 22998095460006381 **|f** 1 **|c** GLRF **|m** BOOK **|a** 39015062862704 **|e** GLRF **|u** mdp.39015062862704 ic 20201013 **|7** 23998095390006381 **|z** v.7 **|p** 08 **|r** 2006-02-01 10:59:00 **|h** P 29 .E481 2006 **|d** HATCH **|b** HATCH
974			**|8** 22998095460006381 **|f** 1 **|c** GLRF **|m** BOOK **|a** 39015062862720 **|e** GLRF **|u** mdp.39015062862720 ic 20201013 **|7** 23998095370006381 **|z** v.9 **|p** 08 **|r** 2006-02-01 10:59:00 **|h** P 29 .E481 2006 **|d** HATCH **|b** HATCH
974			**|8** 22998095460006381 **|f** 1 **|c** GLRF **|m** BOOK **|a** 39015062862738 **|e** GLRF **|u** mdp.39015062862738 ic 20201013 **|7** 23998095360006381 **|z** v.10 **|p** 08 **|r** 2006-02-01 10:59:00 **|h** P 29 .E481 2006 **|d** HATCH **|b** HATCH

Ask a Librarian

974	|8 22998095460006381 |f 1 |c GLRF |m BOOK |a 39015062862712 |e GLRF |u mdp.39015062862712 ic 20201013 |7 23998095380006381 |z v.8 |p 08 |r 2006-02-01 10:59:00 |h P 29 .E481 2006 |d HATCH |b HATCH
974	|8 22998095460006381 |f 1 |c GLRF |m BOOK |a 39015062862746 |e GLRF |u mdp.39015062862746 ic 20201013 |7 23998095350006381 |z v.11 |p 08 |r 2006-02-01 10:59:00 |h P 29 .E481 2006 |d HATCH |b HATCH
974	|8 22998095460006381 |f 1 |c GLRF |m BOOK |a 39015062862761 |e GLRF |u mdp.39015062862761 ic 20201013 |7 23998095330006381 |z v.13 |p 08 |r 2006-02-01 10:59:00 |h P 29 .E481 2006 |d HATCH |b HATCH
974	|8 22998095460006381 |f 1 |c GLRF |m BOOK |a 39015062862779 |e GLRF |u mdp.39015062862779 ic 20201013 |7 23998095320006381 |z v.14 |p 08 |r 2006-02-01 10:59:00 |h P 29 .E481 2006 |d HATCH |b HATCH
974	|8 22998095460006381 |f 1 |c GLRF |m BOOK |a 39015062862753 |e GLRF |u mdp.39015062862753 ic 20201013 |7 23998095340006381 |z v.12 |p 08 |r 2006-02-01 10:59:00 |h P 29 .E481 2006 |d HATCH |b HATCH
974	|8 22998095460006381 |f 1 |c GLRF |m BOOK |a 39015062862977 |e GLRF |u mdp.39015062862977 ic 20201013 |7 23998095420006381 |z v.4 |p 08 |r 2006-02-01 10:59:00 |h P 29 .E481 2006 |d HATCH |b HATCH
974	|8 22998095460006381 |f 1 |c GLRF |m BOOK |a 39015062862993 |e GLRF |u mdp.39015062862993 ic 20201013 |7 23998095400006381 |z v.6 |p 08 |r 2006-02-01 10:59:00 |h P 29 .E481 2006 |d HATCH |b HATCH
974	|8 22998095460006381 |f 1 |c GLRF |m BOOK |a 39015062862985 |e GLRF |u mdp.39015062862985 ic 20210228 |7 23998095410006381 |z v.5 |p 08 |r 2006-02-01 10:59:00 |h P 29 .E481 2006 |d HATCH |b HATCH
974	|8 22998095460006381 |f 1 |c GLRF |m BOOK |a 39015062862951 |e GLRF |u mdp.39015062862951 ic 20201013 |7 23998095440006381 |z v.2 |p 08 |r 2006-02-01 10:59:00 |h P 29 .E481 2006 |d HATCH |b HATCH
974	|8 22998095460006381 |f 1 |c GLRF |m BOOK |a 39015062862969 |e GLRF |u mdp.39015062862969 ic 20201013 |7 23998095430006381 |z v.3 |p 08 |r 2006-02-01 10:59:00 |h P 29 .E481 2006 |d HATCH |b HATCH
974	|8 22998095460006381 |f 1 |c GLRF |m BOOK |a 39015062862944 |e GLRF |u mdp.39015062862944 ic 20201013 |7 23998095450006381 |z v.1 |p 08 |r 2005-12-20 10:59:00 |h P 29 .E481 2006 |d HATCH |b HATCH

Give feedback about this page

Home

Accessibility

Ask a Librarian

Tips for Using Library Search

Get research help

About Library Search

Make an I.L.L. Request

©2022 Regents of the University of Michigan. For details and exceptions, see the Copyright Policy.

Ask a Librarian

ATTACHMENT

D

https://web.archive.org/web/20050319025600
ScienceDire
http://www.dragon-medical-transcription.com/history_speech_recognition.html
Go
FEB MAR JUL
19
2004 2005 2006
261 captures
19 Mar 2005 - 31 Dec 2021
About this capture

History of Speech & Voice Recognition and Transcription Software

Dragon Naturally Speaking Medical Transcription Software
Dragon Medical
Dragon Professional
Dragon Preferred
Transcription Software Comparisons
System Requirements
Dragon Hardware Comp. List
Feedback
Dragon Tutorial
HIPAA Statement

Dragon Add-on Vocabularies
Trigram



Please also view the Speech Recognition Timeline

> SPEECH & VOICE RECOGNITION:

The ability of machines to respond to spoken commands. Speech & Voice recognition enables "hands-free" control of various electronic devices—a particular boon to many disabled persons—and the automatic creation of "print-ready" dictation. Among the earliest applications for speech & voice recognition were automated telephone systems and medical dictation software (Transcription).

"Speech Recognition." Encyclopædia Britannica. 2003. Encyclopædia Britannica Premium Service. 08 Oct, 2003 <http://www.britannica.com/eb/article?eu=139072>.

The technology of Automatic Speech (Voice)Recognition (ASR) and Transcription has progressed greatly over the past few years. Ever since research of this technology began in 1936, the largest barriers to the speed and accuracy of speech & voice recognition was computer speed and power (or lack there of). With the average the CPU now at and above a Pentium III and RAM levels at 500 MB and up, **accuracy levels have reached 95% and better with transcription speeds at over 160 words per minutes.**

As mentioned above, the study of automatic speech recognition and transcription began in the 1936 with ATT&T's Bell Labs. At that time, most research was funded and performed by Universities and the U.S. Government (primarily by the Military and DARPA - Defense Advanced Research Project Agency). It was not until the early 1980's when the technology reached the commercial market.

The Wayback Machine - https://web.archive.org/web/20050319025603/http://www.dragon-medical-transcription.com:80/history_speech_recogn…

History of Speech & Voice Recognition and Transcription Software

Dragon Naturally Speaking Medical Transcription Software
Dragon Medical
Dragon Professional
Dragon Preferred
Transcription Software Comparisons
System Requirements
Dragon Hardware Comp. List
Feedback
Dragon Tutorial
HIPAA Statement

Dragon Add-on Vocabularies
Trigram



Please also view the Speech Recognition Timeline

> SPEECH & VOICE RECOGNITION:

The ability of machines to respond to spoken commands. Speech & Voice recognition enables “hands-free” control of various electronic devices—a particular boon to many disabled persons—and the automatic creation of “print-ready” dictation. Among the earliest applications for speech & voice recognition were automated telephone systems and medical dictation software (Transcription).

"Speech Recognition." Encyclopædia Britannica. 2003. Encyclopædia Britannica Premium Service. 08 Oct, 2003 <http://www.britannica.com/eb/article?eu=139072>.

The technology of Automatic Speech (Voice)Recognition (ASR) and Transcription has progressed greatly over the past few years. Ever since research of this technology began in 1936, the largest barriers to the speed and accuracy of speech & voice recognition was computer speed and power (or lack there of). With the average the CPU now at and above a Pentium III and RAM levels at 500 MB and up, **accuracy levels have reached 95% and better with transcription speeds at over 160 words per minutes.**

As mentioned above, the study of automatic speech recognition and transcription began in the 1936 with ATT&T's Bell Labs. At that time, most research was funded and performed by Universities and the U.S. Government (primarily by the Military and DARPA - Defense Advanced Research Project Agency). It was not until the early 1980's when the technology reached the commercial market.

Like most emerging technologies, there were several competing research "camps", each working independently to develop speech recognition. Please view the Speech Recognition Timeline to get a full view of its development.

The first company to launch a commercial product was Covox in 1982. Covox brought digital sound

IPR2022-01289 Page 00061

HIPAA Made Easy Compliance Software
Main Page
Physicians General
Employer's Medical

Technology History
Speech Recognition

Descriptions
Transcription
Medical Transcription
Voice Recognition
Voice Recognition Software
Dragon Naturally Speaking

(via The Voice Master, Sound Master and The Speech Thing) to the Commodore 64, Atari 400/800, and finally to the IBM PC in the mid '80s. Along with (or bundled) this introduction of sound to computers came Speech Recognition.

Another company that was founded in 1982 and whose eventual product has become the overwhelming leader in the speech recognition market was **Dragon Systems**. Scansoft, Inc. now owns and manufactures this product, Dragon Naturally Speaking.

Dragon Systems History

[1] Dragon Systems was founded in 1982 by James and Janet Baker to commercialize speech recognition technology. As graduate students at Rockefeller University in 1970, they became interested in speech recognition while observing waveforms of speech on an oscilloscope. At the time, systems were in place for recognizing a few hundred words of discrete speech, provided the system was trained on the speaker and the speaker paused between words. There were not yet techniques that could sort through naturally spoken sentences. James Baker saw the waveforms--and the problem of natural speech recognition--as an interesting pattern-recognition problem.

Rockefeller had neither experts in speech understanding nor suitable computing power, and so the Bakers moved to Carnegie Mellon University (CMU), a prime contractor for DARPA's Speech Understanding Research program. There they began to work on natural speech recognition capabilities. Their approach differed from that of other speech researchers, most of whom were attempting to recognize spoken language by providing contextual information, such as the speaker's identity, what the speaker knew, and what the speaker might be trying to say, in addition to rules of English. The Bakers' approach was based purely on statistical relationships, such as the probability that any two or three words would appear one after another in spoken English. They created a phonetic dictionary with the sounds of different word groups and then set to work on an algorithm to decipher a string of spoken words based on phonetic sound matches and the probability that someone would speak the words in that order. Their approach soon began outperforming competing systems.

After receiving their doctorates from CMU in 1975, the Bakers joined IBM's T.J. Watson Research Center, one of the only organizations at the time working on large-vocabulary, continuous speech recognition. The Bakers developed a program that could recognize speech from a 1,000-word vocabulary, but it could not do so in real time. Running on an IBM System 370 computer, it took roughly an hour to decode a single spoken sentence. Nevertheless, the Bakers grew impatient with what they saw as IBM's reluctance to develop simpler systems that could be more rapidly put to commercial use. They left in 1979 to join Verbex Voice Systems, a subsidiary of Exxon Enterprises that had built a system for collecting data over the telephone using spoken digits. Less than 3 years later, however, Exxon exited the speech recognition

IPR2022-01289 Page 00062

business.

With few alternatives, the Bakers decided to start their own company, Dragon Systems. The company survived its early years through a mix of custom projects, government research contracts, and new products that relied on the more mature discrete speech recognition technology. In 1984, they provided Apricot Computer, a British company, with the first speech recognition capability for a personal computer (PC). It allowed users to open files and run programs using spoken commands. But Apricot folded shortly thereafter. In 1986, Dragon Systems was awarded the first of a series of contracts from DARPA to advance large-vocabulary, speaker-independent continuous speech recognition, and by 1988, Dragon conducted the first public demonstration of a PC-based discrete speech recognition system, boasting an 8,000-word vocabulary.

In 1990, Dragon demonstrated a 5,000-word continuous speech system for PCs and introduced DragonDictate 30K, the first large-vocabulary, speech-to-text system for general-purpose dictation. It allowed control of a PC using voice commands only and found acceptance among the disabled. The system had limited appeal in the broader marketplace because it required users to pause between words. Other federal contracts enabled Dragon to improve its technology. In 1991, Dragon received a contract from DARPA for work on machine-assisted translation systems, and in 1993, Dragon received a federal Technology Reinvestment Project award to develop, in collaboration with Analog Devices Corporation, continuous speech & voice recognition systems for desktop and hand-held personal digital assistants (PDAs). Dragon demonstrated PDA speech recognition in the Apple Newton MessagePad 2000 in 1997.

Late in 1993, the Bakers realized that improvements in desktop computers would soon allow continuous voice recognition. They quickly began setting up a new development team to build such a product. To finance the needed expansion of its engineering, marketing, and sales staff, Dragon brokered a deal whereby Seagate Technologies bought 25 percent of Dragon's stock. By July 1997, Dragon had launched Dragon NaturallySpeaking, a continuous speech & voice recognition program for general-purpose use with a vocabulary of 23,000 words. The package won rave reviews and numerous awards. IBM quickly followed suit, offering its own continuous speech recognition program, ViaVoice, in August after a crash development program. By the end of the year, the two companies combined had sold more than 75,000 copies of their software. Other companies, such as Microsoft Corporation and Lucent Technologies, are expected to introduce products in the near future, and analysts expect a $4 billion worldwide market by 2001.

In 2000, Lernout & Hauspie acquired Dragon Systems. In 2001, Scansoft, Inc. acquired all rights to Lernout & Hauspie's speech recognition products including Dragon Naturally Speaking. In 2003, Scansoft, Inc. acquires Speechworks.

Scansoft, Inc. is presently the world leader in the technology of Speech Recognition in the commercial market.

IPR2022-01289 Page 00063

[1] Funding a Revolution:Government Support for Computing Research . Copyright 1999 by the National Academy of Sciences. http://www.nap.edu/readingroom/books/far/ch9_b2.html

SOURCE: The primary source for this history is Garfinkel (1998).

About Us Contact Us Site Map
Dragon Naturally Speaking Medical Transcription
Voice Recognition Voice Recognition Software Transcription

IPR2022-01289 Page 00064

ATTACHMENT

E

BOOK

Handbook of natural language processing

Full Record MARC Tags

Main title

Handbook of natural language processing / edited by Nitin Indurkhya, Fred J. Damerau.

Published/Created

Boca Raton, FL : Chapman & Hall/CRC, c2010.

Request this Item LC Find It

More Information

LCCN Permalink	https://lccn.loc.gov/2009049507
Description	xxiii, 678 p. : ill. ; 26 cm.
ISBN	9781420085921 (alk. paper) 1420085921 (alk. paper)
LC classification	QA76.9.N38 H363 2010
Related names	Indurkhya, Nitin. Damerau, Frederick J. (Frederick Jacob), 1931-
LC Subjects	Natural language processing (Computer science)--Handbooks, manuals, etc.

Browse by shelf order	QA76.9.N38
Notes	Includes bibliographical references and index.
Series	Chapman & Hall/CRC machine learning & pattern recognition series Chapman & Hall/CRC machine learning & pattern recognition series.
LCCN	2009049507
Dewey class no.	006.3/5
Other system no.	(OCoLC)ocn213318186
Type of material	Book

Item Availability

CALL NUMBER	QA76.9.N38 H363 2010 Copy 2
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged

CALL NUMBER	QA76.9.N38 H363 2010 CABIN BRANCH Copy 1
Request in	Jefferson or Adams Building Reading Rooms - STORED OFFSITE
Status	c.1 Charged - Due - (Internal Loan)

ATTACHMENT

F

MARC bibliographic record for *Handbook of Natural Language Processing*, including the Huang and Deng chapter, in the shared online library catalog of the Indiana University system. Available at: https://iucat.iu.edu/catalog/9533311

Librarian View

LEADER 00000cam a2200000 a 45 0		
001 9533311		
003 OCoLC		
005 20101014165528.1		
008 091123s2010 flua f 001 0 eng||		
010		a| 2009049507
035		a| (Sirsi)9533311
035		a| (OCoLC)ocn213318186
035		a| (OCoLC)213318186
040		a| DLC c| DLC d| YDX d| YDXCP d| CDX d| BWX d| SNK d| SOI
020		a| 9781420085921 (alk. paper)
020		a| 1420085921 (alk. paper)
050	0 0	a| QA76.9.N38 b| H363 2010
082	0 0	a| 006.3/5 2| 22
049		a| I3UA
245	0 0	a| Handbook of natural language processing / c| edited by Nitin Indurkhya, Fred J. Damerau.
260		a| Boca Raton, FL : b| Chapman & Hall/CRC, c| c2010.
300		a| xxiii, 678 p. : b| ill. ; c| 26 cm.
490	1	a| Chapman & Hall/CRC machine learning & pattern recognition series
504		a| Includes bibliographical references and index.

505	0		a| Classical approaches to natural language processing / Robert Dale -- Text preprocessing / Davis D. Palmer -- Lexical analysis / Andrew Hippisley -- Syntactic parsing / Peter Ljunglöf and Mats Wirén -- Semantic analysis / Cliff Goddard and Andrea C. Schalley -- Natural language generation / Davis D. McDonald -- Corpus creation / Richard Xiao -- Treebank annotation / Eva Hajičová ... [et al.] -- Fundamental statistical techniques / Tong Zhang -- Part-of-speech tagging / Tunga Güngör -- Statistical Parsing / Joakim Nivre -- Multiword expressions / Timothy Baldwin and Su Nam Kim -- Normalized web distance and word similarity / Paul M.B. Vitányi and Rudi L. Cilibrasi -- Word sense disambiguation / Davis Yarowsky -- An overview of modern speech recognition / Xuedong Huang and Li Deng -- Alignment / Dekai Wu -- Statistical machine translation / Abraham Ittycheriah -- Chinese machine translation / Pascale Fung -- Information retrieval / Jacques Savoy and Eric Gaussier -- Question answering / Diego Mollá-Aliod and José-Luis Vicedo -- Information extraction / Jerry R. Hobbs and Ellen Riloff -- Report generation / Leo Wanner -- Emerging applications of natural language generation in information visualization, education, and health care / Barbara Di Eugenio and Nancy L. Green -- Ontology construction / Philipp Cimiano, Johanna Völker, and Paul Buitelaar -- BioNPL: biomedical text mining / K. Bretonnel Cohen -- Sentiment analysis and subjectivity / Bing Liu.
520	1		a| "The Handbook of Natural Language Processing, Second Edition presents practical tools and techniques for implementing natural language processing in computer systems. Along with removing outdated material, this edition updates every chapter and expands the content to include emerging areas, such as sentiment analysis."--Book jacket.
650		0	a| Natural language processing (Computer science) v| Handbooks, manuals, etc. =| ^A915756
700	1		a| Indurkhya, Nitin. =| ^A1468836
700	1		a| Damerau, Frederick J. q| (Frederick Jacob), d| 1931-2009 =| ^A2838850
830		0	a| Chapman & Hall/CRC machine learning & pattern recognition series. =| ^A2827484
856	4	2	3| Companion wiki u| http://handbookofnlp.cse.unsw.edu.au
596			a| 37 42 46
999			a| QA76.9.N38 H363 2010 i| 30000127555351 j| QA 000076.9 .N38 H363 2010 l| STACKS m| I-UNIVLIB n| STACKS w| LC z| CAT;05-12-2011;UL009
999			a| QA76.9.N38 H363 2010 i| 30000127676777 j| QA 000076.9 .N38 H363 2010 l| STACKS m| SBEND n| STACKS w| LC z| CAT;04-04-2011;SB429

ATTACHMENT

G

```
LEADER 00000cam  2200000 a 4500
001    213318186
003    OCoLC
005    20100416085517.0
008    091123s2010    flua     bf   001 0 eng
010    2009049507
020    9781420085921 (alk. paper)
020    1420085921 (alk. paper)
035    (OCoLC)213318186
040    DLC|cDLC|dYDX|dYDXCP|dCDX|dBWX
040    |dOrLoB-B
049    WTUU
050 00 QA76.9.N38|bH363 2010
050 00 QA76.9.N38|bH363 2010
082 00 006.3/5|222
245 00 Handbook of natural language processing /|cedited by Nitin
       Indurkhya, Fred J. Damerau.
260    Boca Raton, FL :|bChapman & Hall/CRC,|cc2010.
300    xxiii, 678 p. :|bill. ;|c26 cm.
490 1  Chapman & Hall/CRC machine learning & pattern recognition
       series
504    Includes bibliographical references and index.
520 1  "The Handbook of Natural Language Processing, Second
       Edition presents practical tools and techniques for
       implementing natural language processing in computer
       systems. Along with removing outdated material, this
       edition updates every chapter and expands the content to
       include emerging areas, such as sentiment analysis."--BOOK
       JACKET.
650  0 Natural language processing (Computer science)|vHandbooks,
       manuals, etc.
700 1  Indurkhya, Nitin.
700 1  Damerau, Frederick J.|q(Frederick Jacob),|d1931-2009
830  0 Chapman & Hall/CRC machine learning & pattern recognition
       series.
910    sw
938    YBP Library Services|bYANK|n3118893
938    Coutts Information Services|bCOUT|n7640164|c63.99 GBP
938    Blackwell Book Service|bBBUS|nR7835470|c$99.95
970 01 |tList of Figures
970 01 |tList of Tables
970 11 |lPt. I|tClassical Approaches
970 21 |l1|tClassical Approaches to Natural Language Processing
       |cRobert Dale|fDale, Robert
970 21 |l2|tText Preprocessing|cDavid D. Palmer|fPalmer, David D.
970 21 |l3|tLexical Analysis|cAndrew Hippisley|fHippisley, Andrew
970 21 |l4|tSyntactic Parsing|cPeter Ljunglof|fLjunglof, Peter
       |cMats Wiren|fWiren, Mats
```

```
970 21 |l5|tSemantic Analysis|cCliff Goddard|fGoddard, Cliff
       |cAndrea C. Schalley|fSchalley, Andrea C.
970 21 |l6|tNatural Language Generation|cDavid D. McDonald
       fMcDonald, David D.
970 11 |lPt. II|tEmpirical and Statistical Approaches
970 21 |l7|tCorpus Creation|cRichard Xiao|fXiao, Richard
970 21 |l8|tTreebank Annotation|cEva Hajicova|fHajicova, Eva
       |cAnne Abeille|fAbeille, Anne|cJan Hajic|fHajic, Jan|cJiri
       Mirovsky|fMirovsky, Jiri|cZdenka Uresova|fUresova, Zdenka
970 21 |l9|tFundamental Statistical Techniques|cTong Zhang|fTong,
       Zhang
970 21 |l10|tPart-of-Speech Tagging|cTunga Gungor|fGungor, Tunga
970 21 |l11|tStatistical Parsing|cJoakim Nivre|fNivre, Joakim
970 21 |l12|tMultiword Expressions|cTimothy Baldwin|fBaldwin,
       Timothy|cSu Nam Kim|fSu, Nam Kim
970 21 |l13|tNormalized Web Distance and Word Similarity|cPaul M.
       B. Vitanyi|fVitanyi, Paul M. B.|cRudi L. Cilibrasi
       |fCilibrasi, Rudi L.
970 21 |l14|tWord-Sense Disambiguation|cDavid Yarowsky|fYarowsky,
       David
970 21 |l15|tAn Overview of Modern Speech Recognition|cXuedong
       Huang|fHuang, Xuedong|cLi Deng|fLi, Deng
970 21 |l16|tAlignment|cDekai Wu|fWu, Dekai
970 21 |l17|tStatistical Machine Translation|cAbraham Ittycheriah
       fIttycheriah, Abraham
970 11 |lPt. III|tApplications
970 21 |l18|tChinese Machine Translation|cPascale Fung|fFung,
       Pascale
970 21 |l19|tInformation Retrieval|cJacques Savoy|fSavoy, Jacques
       |cEric Gaussier|fGaussier, Eric
970 21 |l20|tQuestion Answering|cDiego Molla-Aliod|fMolla-Aliod,
       Diego|cJose-Luis Vicedo|fVicedo, Jose-Luis
970 21 |l21|tInformation Extraction|cJerry R. Hobbs|fHobbs, Jerry
       R.|cEllen Riloff|fRiloff, Ellen
970 21 |l22|tReport Generation|cLeo Wanner|fWanner, Leo
970 21 |l23|tEmerging Applications of Natural Language Generation
       in Information Visualization, Education, and Healthcare
       |cBarbara Di Eugenio|fDi Eugenio, Barbara|cNancy L. Green
       fGreen, Nancy L.
970 21 |l24|tOntology Construction|cPhilipp Cimiano|fCimiano,
       Philipp|cJohanna Volker|fVolker, Johanna|cPaul Buitelaar
       |fBuitelaar, Paul
970 21 |l25|tBioNLP: Biomedical Text Mining|cK. Bretonnel Cohen
       |fCohen, K. Bretonnel
970 21 |l26|tSentiment Analysis and Subjectivity|cBing Liu|fBing,
       Liu
970 01 |tIndex
994    C0|bWTU
```

LOCATION	CALL #	STATUS
Olin Library Level B Stacks	QA76.9.N38 H363 2010 *Browse Nearby Call Numbers*	NOT CHECKED OUT

Ask Us! 24/7 Chat - Danforth Campus

Medical Campus Patrons contact Becker Medical Library Services

IPR2022-01289 Page 00073

For Students	Ask Us
For Faculty	Research Support
For Staff	Staff Directory
Visitors & Alumni	Make a Gift

University Libraries
MSC 1061-141-B
Washington University in St. Louis
One Brookings Dr.
St. Louis, MO 63130

IPR2022-01289 Page 00074

ATTACHMENT

H

BOOK

Handbook of natural language processing

Full Record MARC Tags

000		01264cam a2200289 a 4500
001		15993377
005		20100602131311.0
008		091123s2010 flua bf 001 0 eng
906	__	|**a** 7 |**b** cbc |**c** orignew |**d** 1 |**e** ecip |**f** 20 |**g** y-gencatlg
925	0_	|**a** acquire |**b** 2 shelf copies |**x** policy default
955	__	|**b** xh11 2009-11-23 |**i** xh11 2009-11-23 to Dewey |**w** rd14 2009-11-23 |**f** rc09 2010-06-02 Z-CipVer |**t** rc09 2010-06-02 c. 2 added
010	__	|**a** 2009049507
020	__	|**a** 9781420085921 (alk. paper)
020	__	|**a** 1420085921 (alk. paper)
035	__	|**a** (OCoLC)ocn213318186
040	__	|**a** DLC |**c** DLC |**d** YDX |**d** YDXCP |**d** CDX |**d** BWX |**d** DLC
050	00	|**a** QA76.9.N38 |**b** H363 2010
082	00	|**a** 006.3/5 |**2** 22
245	00	|**a** Handbook of natural language processing / |**c** edited by Nitin Indurkhya, Fred J. Damerau.
260	__	|**a** Boca Raton, FL : |**b** Chapman & Hall/CRC, |**c** c2010.
300	__	|**a** xxiii, 678 p. : |**b** ill. ; |**c** 26 cm.
490	1_	|**a** Chapman & Hall/CRC machine learning & pattern recognition series
504	__	|**a** Includes bibliographical references and index.
650	_0	|**a** Natural language processing (Computer science) |**v** Handbooks, manuals, etc.
700	1_	|**a** Indurkhya, Nitin.
700	1_	|**a** Damerau, Frederick J. |**q** (Frederick Jacob), |**d** 1931-
830	_0	|**a** Chapman & Hall/CRC machine learning & pattern recognition series.

Request this Item | LC Find It

Item Availability

CALL NUMBER	QA76.9.N38 H363 2010 Copy 2
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged

CALL NUMBER	QA76.9.N38 H363 2010 CABIN BRANCH Copy 1
Request in	Jefferson or Adams Building Reading Rooms - STORED OFFSITE
Status	c.1 Charged - Due - (Internal Loan)

ATTACHMENT

I

Context Sensitive Links Export Add To Marked List

1 of 1

An Overview of Modern Speech Recognition

By: Huang, XD (Huang, Xuedong) [1] ; Deng, L (Deng, Li) [1]
Edited by: Indurkhya, N (Indurkhya, N) ; Damerau, FJ (Damerau, FJ)
HANDBOOK OF NATURAL LANGUAGE PROCESSING, SECOND EDITION
Book Series: Chapman & Hall-CRC Machine Learning & Pattern Recognition Series
Page: 339-366
Published: 2010
Indexed: 2010-01-01
Document Type: Article; Book Chapter

Keywords

Keywords Plus: MAXIMUM-LIKELIHOOD; REGRESSION; HMMS

Author Information

Corresponding Address: Huang, Xuedong (corresponding author)
Microsoft Corp, Microsoft Res, Redmond, WA 98052 USA
Addresses:
1 Microsoft Corp, Microsoft Res, Redmond, WA 98052 USA

Categories/Classification

Research Areas: Computer Science; Linguistics

+ See more data fields

Journal information

HANDBOOK OF NATURAL LANGUAGE PROCESSING, SECOND EDITION
Current Publisher: CRC PRESS-TAYLOR & FRANCIS GROUP, 6000 BROKEN SOUND PARKWAY NW, STE 300, BOCA RATON, FL 33487-2742 USA
Research Areas: Computer Science; Linguistics
Web of Science Categories: Computer Science, Artificial Intelligence; Computer Science, Interdisciplinary Applications; Linguistics

Citation Network

In Web of Science Core Collection

20
Citations

Create citation alert

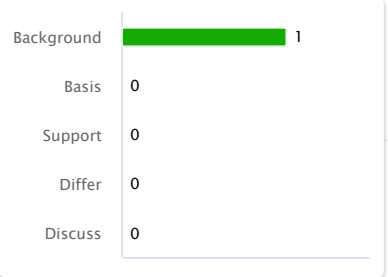
20
Times Cited in All Databases

67
Cited References
View Related Records

+ See more times cited

Citing items by classification New

Breakdown of how this article has been mentioned, based on available citation context data and snippets from 1 citing item(s).



Most Recently Cited by

Kaur, G; Srivastava, M; Kumar, A;
Speech Recognition Using Enhanced Features with Deep Belief Network for Real Time Application
WIRELESS PERSONAL COMMUNICATIONS

Bhatt, S; Jain, A; Dev, A;
Acoustic Modeling in Speech Recognition: Systematic Review
INTERNATIONAL JOURNAL OF ADVANCED

44

IPR2022-01289 Page 00080

ATTACHMENT

J


https://web.archive.org/web/20210423171339/https://www.cs.rochester.edu/u/james/CSC248/Lec12.p
Search
INTERNET ARCHIVE
WaybackMachine
https://www.cs.rochester.edu/u/james/CSC248/Lec12.pdf
Go
24 captures
4 Nov 2003 - 23 Apr 2021
NOV
APR
23
MAY
2019
2021
2022
About this capture

Lecture 12: An Overview of Speech Recognition

1. Introduction

We can classify speech recognition tasks and systems along a set of dimensions that produce various tradeoffs in applicability and robustness.

Isolated word versus continuous speech: Some speech systems only need identify single words at a time (e.g., speaking a number to route a phone call to a company to the appropriate person), while others must recognize sequences of words at a time. The isolated word systems are, not surprisingly, easier to construct and can be quite robust as they have a complete set of patterns for the possible inputs. Continuous word systems cannot have complete representations of all possible inputs, but must assemble patterns of smaller speech events (e.g., words) into larger sequences (e.g., sentences).

Speaker dependent versus speaker independent systems: A speaker dependent system is a system where the speech patterns are constructed (or adapted) to a single speaker. Speaker independent systems must handle a wide range of speakers. Speaker dependent systems are more accurate, but the training is not feasible in many applications. For instance, an automated telephone operator system must handle any person that calls in, and cannot ask the person to go through a training phase before using the system. With a dictation system on your personal computer, on the other hand, it is feasible to ask the user to perform a hour or so of training in order to build a recognition model.

Small versus vocabulary systems: Small vocabulary systems are typically less than 100 words (e.g., a speech interface for long distance dialing), and it is possible to get quite accurate recognition for a wide range of users. Large vocabulary systems (e.g., say 20,000 words or greater), typically need to be speaker dependent to get good accuracy (at least for systems that recognize in real time). Finally, there are mid-size systems, on the order to 1000-3000 words, which are typical sizes for current research-based spoken dialogue systems.

Some applications can make every restrictive assumption possible. For instance, voice dialing on cell phones has a small vocabulary (less than 100 names), is speaker dependent (the user says every word that needs to be recognized a couple of times to train it), and isolated word. On the other extreme, there are research systems that attempt to transcribe recordings of meetings among several people. These must handle speaker independent, continuous speech, with large vocabularies. At present, the best research systems cannot achieve much better than a 50% recognition rate, even with fairly high quality recordings.

2. Speech Recognition as a "Tagging" Problem,

Speech recognition can be viewed as a generalization of the tagging problem: Given an acoustic output $A_{1,T}$ (consisting of a sequence of acoustic events a_1, …, a_T) we want to find a sequence of words $W_{1,R}$ that maximizes the probability

IPR2022-01289 Page 00083

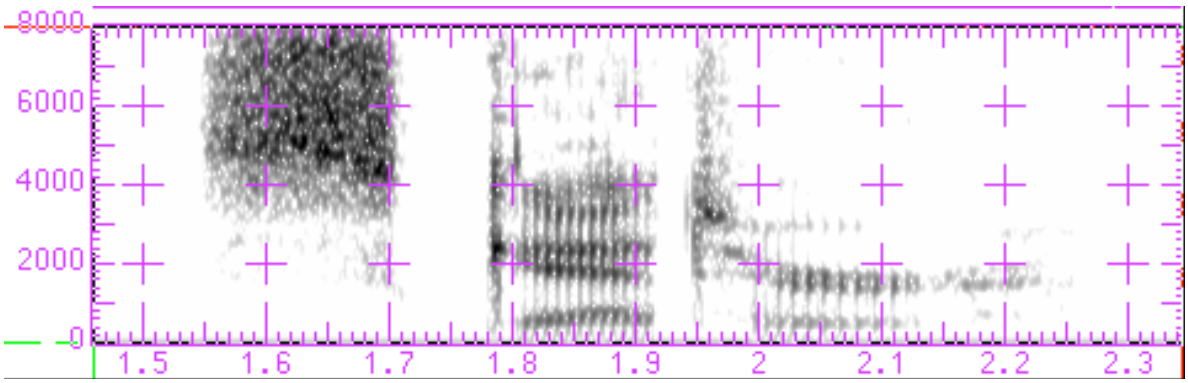


Figure 1: A Spectogram of the word "sad"

$$\text{Argmax}_W \, P(W_{1,R} \mid A_{1,T})$$

Using the standard reformulation for tagging by rewriting this by Bayes Rule and then dropping the denominator since it doesn't affect the maximization, we transform the problem into computing

$$\text{argmax}_W \, P(A_{1,T} \mid W_{1,R}) \, P(W_{1,R})$$

In the speech recognition work, $P(W_{1,R})$ is called the **language model** as before, and $P(A_{1,T} \mid W_{1,R})$ is called the **acoustic model**.

This formulation so far, however, seems to raise more questions that answers. In particular, we will address the main issues briefly here and then return to look at them in detail in the following chapters.

What does a speech signal look like?

Human speech can be best explored by looking at the intensity at different frequencies over time. This can be shown graphical in a **spectogram** of a signal containing one word such as the one shown in Figure 1. Time is on the X axis, frequency is on the Y axis, and the darkness of the area corresponds to intensity. This word starts out with a lot of high frequency noise with no noticeable lower frequencies or resonance (typical of an /s/ or /sh/ sound), starting at time 1.55, then at 1.7 there is a period of silence. This initial signal is indicative of a "stop" consonant such as a /t/, /d/, /p/. Between 1.8 and 1.9 we see a typical vowel, with strong lower frequencies and distinctive bars of resonance called the **formants** clearly visible. After 1.9 it looks like another stop consonant, that includes an area with lower frequency and resonance.

The Sounds of Language

The sounds of language are classified into what are called **phonemes**. A phoneme is minimal unit of sound that has semantic content. e.g., the phoneme AE versus the phoneme EH captures the difference between the words "bat" and "bet". Not all acoustic

IPR2022-01289 Page 00084

changes change meaning. For instance, singing words at different notes doesn't change meaning in English. Thus changes in pitch does not lead to phenemic distinctions.

Often we talk of specific features by which the phonemes can be distinguished. One of the most important features distinguishing phonemes is **Voicing**. A voiced phoneme is one that involves sound from the vocal chords. For instance, F (e.g., "fat") and V (e.g., "vat") differ primarily in the fact that in the latter the vocal chords are producing sound, (i.e., is **voiced)**, whereas the former does not (ie.., it is **unvoiced**).

Here's a quick breakdown of the major classes of phonemes.

Vowels

Vowels are always voiced, and the differences in English depend mostly on the formants (prominent resonances) that are produced by different positions of the tongue and lips. Vowels generally stay steady over the time they are produced. By holding the tongue to the front and varying its height we get vowels IY (beat), IH (bit), EH (bat), and AE (bet). With tongue in the mid position, we get AA (Bob - Bahb), ER (Bird), AH (but), and AO (bought). With the tongue in the Back Position we get UW (boot), UH (book), and OW (boat). There is another class of vowels called the **Diphongs,** which do change during their duration. They can be thought of as starting with one vowel and ending with another. For example, AY (buy) can be approximated by starting with AA (Bob) and ending with IY - (beat). Similarily, we have AW (down, cf. AA W), EY (bait, cf. EH IY), and OY (boy, cf. AO IY).

Consonants

Consonants fall into general classes, with many classes having voiced and unvoiced members.

Stops or Plosives: These all involve stopping the speech stream (using the lips, tongue, and so on) and then rapidly releasing a stream of sound. They come in unvoiced, voiced pairs: P and B, T and D, and K and G.

Fricatives: These all involves "hissing" sounds generated by constraining the speech stream by the lips, teeth and so on. They also come in unvoiced, voiced pairs: F and V, TH and DH (e.g., thing versus that), S and Z, and finally SH and ZH (e.g., shut and azure).

Nasals: These are all voiced and involve moving air through the nasal cavities by blocking it with the lips, gums, and so on. We have M, N, and NX (sing).

Affricatives: These are like stops followed by a fricative: CH (church), JH (judge)

Semi Vowels: These are consonants that have vowel-like characteristics, and include W, Y, L, and R.

Whisper: H

This is the complete set of phonemes typically identified for English.

IPR2022-01289 Page 00085

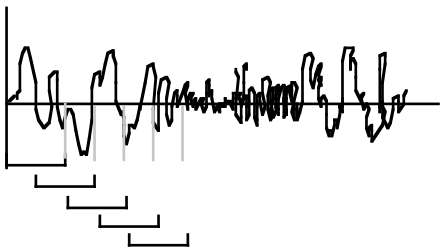


Figure 2: Taking 20 ms segments of speech at 10ms intervals

How do we map from a continuous speech signal to a discrete sequence $A_{1,T}$?

One of the common approaches to mapping a signal to discrete events is to define a set of symbols that correspond to useful acoustic features of the signal over a short time interval. We might first think that phonemes are the ideal unit, but they turn out to be far too complex and extended over time to be classified with simple signal processing techniques. Rather, we need to consider much smaller intervals of speech (and then, as we see in a minute, model phonemes by HMMs over these values). Typically, the speech signal is divided into very small segments (e.g., 20 millseconds), and these intervals often overlap in time as shown in Figure 1. These segments are then classified into a set of different types, each corresponding to a new symbol in a vocabulary called the **codebook**. Speech systems typically use a vocabulary size between 256 and 1024. Each symbol in the codebook is associated with a vector of acoustic features that defines its prototypical acoustic properties. Rather than trying to define the properties of codebook symbols by hand, clustering algorithms are used on training data to find a set of properties that best distinguish the speech data. To do this, however, we first need to design a representation of the speech signal as a relatively small vector of parameters. These parameters must be chosen so that they capture the important distinguishing characteristics that we use to classify different speech sounds. One of the most commonly used features record the intensity of the signal over various bands of frequencies. The lower bands help distinguish parts of voiced sounds, and the upper help distinguish fricatives. In addition, other features capture the rate of change of intensity from one segment to the next. These "delta" parameters are useful for identifying areas of rapid change in the signal, such as at the points where we have stops and plosives. Details on specific useful representations and the training algorithms to build prototype vectors for codebook symbols will be presented in a later lecture.

IPR2022-01289 Page 00086

How do we map from acoustic events (codebook symbols) to words?

Since there are many more acoustic events than there are words, we need to introduce intermediate structures to represent words. Since words will correspond to codebook symbol sequences, it makes sense to use an HMM for each word. These word HMMs then would need to be combined with the language model to produce an HMM model for sentences. To see this technique, consider a highly simplified HMM model for a word such as "sad", and a bigram language model for a language that contains the words "one", "sad" and "but", as shown in Figure 3. We can construct a sentence HMM by simply inserting the word HMM's in place of each word node in the bigram model. Figure 4 shows an HMM where we have substituted just the word "sag". To complete the process we'd do the same from *bus* and *one*. Note that the probabilities from the bigram model are used now to move between the word networks.

How can we train large vocabularies? Applications: Sub Word Models

When we are working with applications of a few hundred words, it is often feasible to obtain enough training data to train each word model individually. As applications grow into the thousands of words, or tens of thousands, it becomes impractical to obtain the necessary training data. Say we need a corpus that has at least 5 instances of each word in order to train reasonably robust patterns. Then for a ten thousand word vocabulary we would need well over 50,000 training words as the words will not be uniformly distributed. Besides being impractical to collect such data, this method would lose all the generalities that could be exploited between words that contain similar sound patterns, and make it very difficult to design systems that could adapt to a particular speaker. Thus, large-vocabulary speech recognition systems do not use word based models, but rather use **subword models**.

The idea is that we build HMM models for parts of words, which can then be reused in different word models. There is a range of possibilities for what to base the subword units on, including the following:

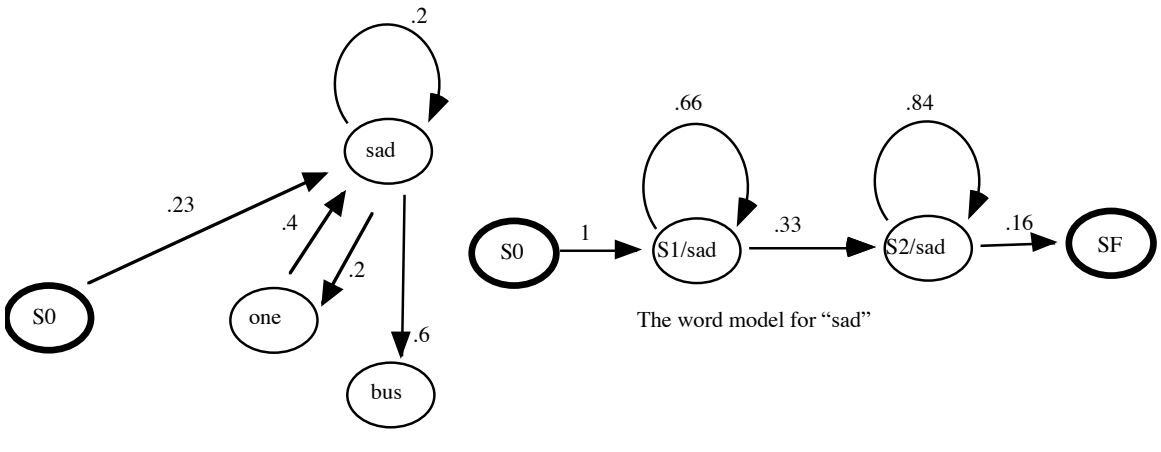


Figure 3: A simple language model and word model

IPR2022-01289 Page 00087

phoneme-based models;

syllable-based models;

demi-syllable based models, where each sound represents a consonant cluster preceding a vowel, or following a vowel;

phonemes in context, or triphones: context dependent models of phonemes dependent on what precedes and follows the phoneme;

There are advantages and disadvantages of each possible subword unit. Phoneme-based models are attractive since there are typically few of them (e.g., we only need 40 to 50 for English). Thus we need relatively little training data to construct good models. Unfortunately, because of co-articulation and other context dependent effects, pure phoneme-based models do not perform well. Some tests have shown they have only half the word accuracy of word based models.

Syllable-based models provide a much more accurate model, but the prospect of having to train the estimated 10,000 syllables in English makes the training problem nearly as difficult as the word based models we had to start with. In a syllable based representation, all single syllable words such as *sad,* would of course be represented identically to the word based models.

Demi-syllables produces a reasonable compromise. Here we'd have models such as /B AE/ and /AE D/ combining to represent the word *bad*. With about 2000 possibilities, training would be difficult but not impossible.

Triphones (or **phonemes in context, PIC**) produce a similar style of representation to demi-syllables. Here, we model each phoneme in terms of its context (i.e., the phones on each side). The word *bad* might break down into the triphone sequence $_{SIL}B_{AE\ B}AE_{D\ AE}D_{SIL,}$ i.e., a B preceded by silence and followed by AE, then AE preceded by B and followed by D, and then D preceded by AE and followed by silence. Of course, in continuous speech, we might not have a silence before of after the word, and we'd then have a different context depending on the words. For instance, the triphone string for the utterance *Bad art* might be $_{SIL}B_{AE\ B}AE_{D\ AE}D_{AA\ D}AA_{R\ AA}R_{T\ R}T_{SIL}$. In other words, the last model in *bad* is a D preceded by AE and followed by AA. Unfortunately, with 50 phonemes, we would have 50^3 (= 125,000!) possible triphones. But many of these don't occur in English. Also we can reduce the number of possibilities by abstracting away the context parameters. For instance, rather than using the 50 phonemes for each context, we use a reduced set including categories like silence, vowel, stop, nasal, liquid consonant, etc (e.g., we have an entry like $_{V}T_{V}$, a /T/ between two vowels, rather than $_{AE}T_{AE}$, $_{AE}T_{AH}$, $_{AE}T_{EH}$, and so on).

Whatever subword units are each, each word in the lexicon is represented by a sequence of units represented as a Finite State Machine or a Markov model. We then have three levels of Markov and Hidden Markov models which need to be combined into a large network. For instance, Figure 4 shows the three levels that might be used for the word *sad* with a phoneme-based model

IPR2022-01289 Page 00088

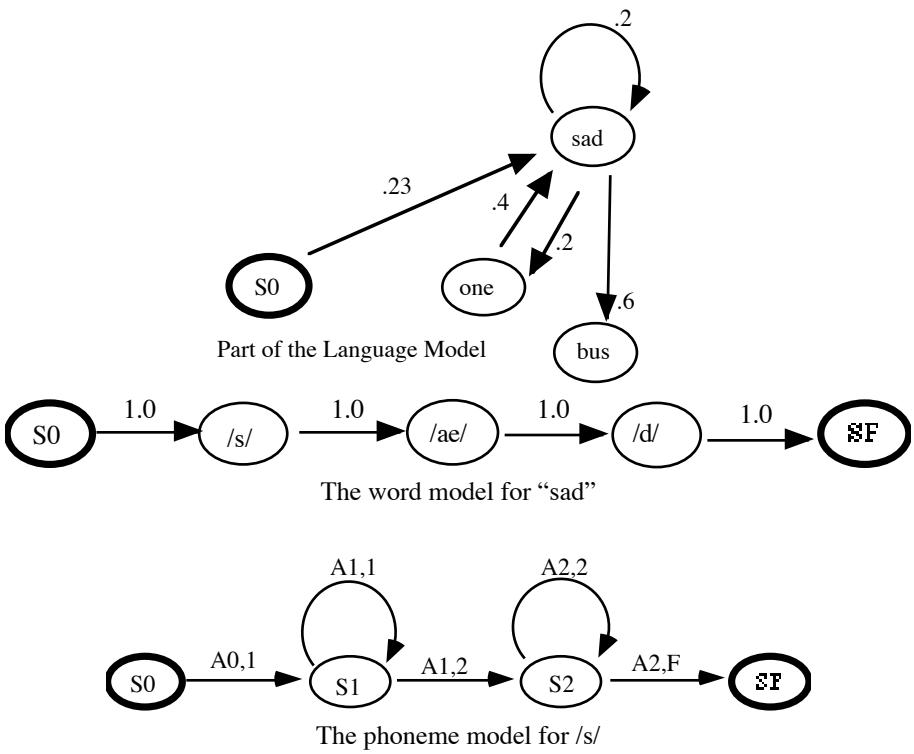


Figure 4: Three Levels in a Phoneme-based model

To construct the exploded HMM, we would first replace the phoneme nodes in the word model for *sad* with the appropriate phoneme models as shown in Figure 5, and then use this network to replace the node for *sad* in the language model as shown in Figure 6. Once this is down for all nodes representing words, we would have single HMM for recognizing utterances using phoneme based pattern recognition.

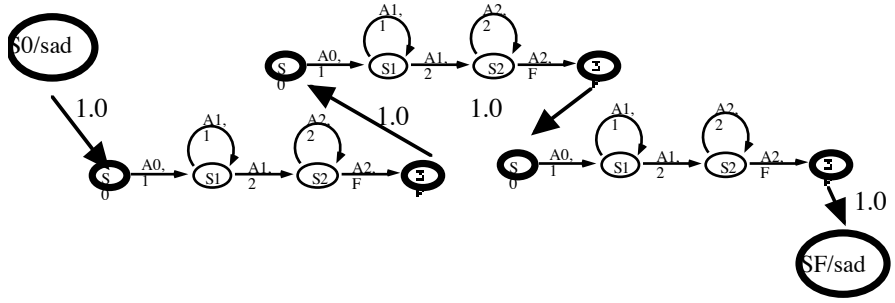


Figure 5: The constructed word model for *sad*

IPR2022-01289 Page 00089

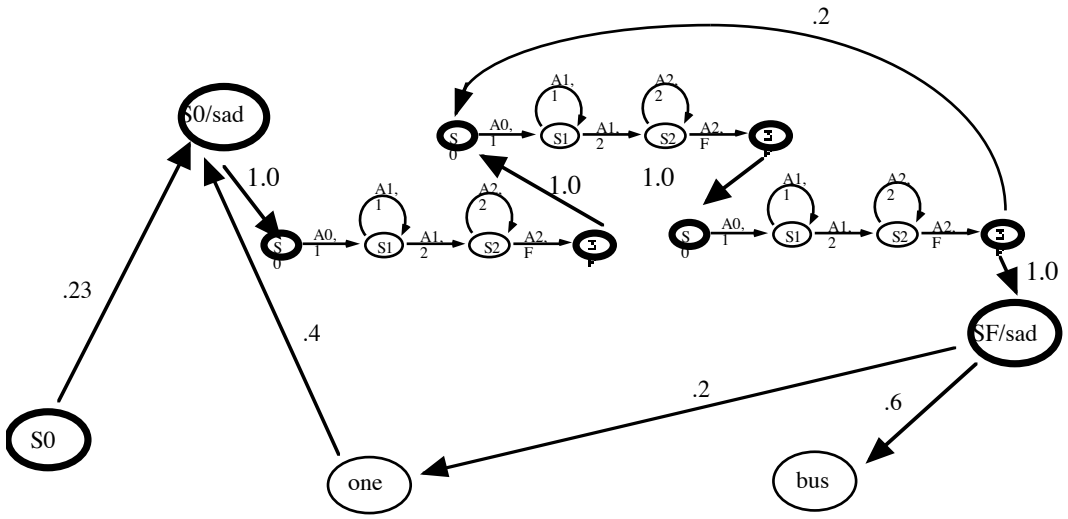


Figure 6: The language model with *sad* expanded

How do we find the word boundaries?

We have solved this problem, in theory at least. If we can run a Viterbi algorithm on a network such as that in Figure 6, the best path will identify where the words are (and in fact provide a phonemic segmentation as well).

How can we train and search such a large network?

While we have already examined methods for searching and training HMMs, the scale of the networks used in speech recognition are such that some additional techniques are needed. For instance, the networks are so large that we need to use a beam or best-first search strategy, and often will build the networks incrementally as the search progresses. This will discussed in a future chapter.

IPR2022-01289 Page 00090

ATTACHMENT

K

CSC 248/448 Home Page

Fall 2003

This pages gives pointers to all the handouts and resources for the course. There is a great website of resources that is a companio and Schutze text.

A description of the course is found in the Course Handout. A syllabus will be available shortly.

Lectures

Lecture 1: Overview and Key Concepts
Lecture 2: A Whirlwind Tour of Probability
Lecture 3: Estimating Probability Distributions
Lecture 4: Estimating with Hold Outs
Lecture 5: More on Estimation
Lecture 6: Entropy and Information Theory
Lecture 7: Basic Automata Models
Lecture 8: Tagging Applications
Lecture 9: Searching HMM Models
Lecture 10: Learning with Hidden Data: The EM Algorithm
Lecture 11: Training HMMs
Lecture 12: Approaches to Speech Recognition
Lecture 13: Signal Processing for Speech-
Lecture 14: Building Code Books
Lecture 15: Search and Training for Large Vocabulary Speech Recognition
Lecture 16: Simple Alignment Models
Lecture 17: Using unordered Contexts
Lecture 18: Information Retrieval
Lecture 19: Probabilistic Context Free Grammars
Lectures 20 and 21: Machine Translation (these are notes from Kevin Knight the first 12 sections review work covered earlier MT starts with section 13)
Lecture 22: Effective CFG Models

Assignments

Assignment 1, Due September 16. There are two test corpora for the assignment. Here are assignment one solutions.
Assignment 2, Due October 2 . The test corpora can be found at
ftp://ftp.cs.rochester.edu/pub/u/myros/multi_corpora.tgz
Here are assignment two solutions.
Assignment 3: Due November 4th. Here is the data file.
Assignment 4: Due November 25th. Here is the training data.
Makeup Assignment, Due December 9th at start of class. No late assignments accepted.

A copy of the midterm from last time is http://www.cs.rochester.edu/u/james/CSC248/oldmidterm.pdf

IPR2022-01289 Page 00092

ATTACHMENT

L

James F. Allen

b. 1950. Ph.D. (1979) University of Toronto. Assistant Professor (79-84), Associate Professor (84-87), Department Chair (87-90), Professor (87-present), John H. Dessauer Professor of Computer Science (92-present), University of Rochester; Senior Research Scientist (02-present), Associate Director (06-present), Florida Institute for Human and Machine Cognition; Editor-in-Chief, Computational Linguistics (83-93; Presidential Young Investigator (84-89); author of Natural Language Understanding, Benjamin Cummings (87), 2nd edition (1995); Reasoning About Plans, Morgan Kaufmann (91); co-editor of Readings in Planning, Morgan Kaufmann (90); Founding Fellow of the AAAI.

My research interests span a range of issues covering natural language understanding, discourse, knowledge representation, common-sense reasoning and planning. I am particularly interested in the overlap between natural language understanding and reasoning. While most of the NLP field has moved to statistical learning methods as the paradigm for language processing, I believe that deep language understanding can only currently be achieved by significant hand-engineering of semantically-rich formalisms coupled with statistical preferences. For further discussion of this viewpoint, see the state of NLP . For a more general discussion of AI, see my keynote address from the 1998 AAAI National conference here . The TRIPS project is a long-term effort to build generic technology for dialogue systems (both spoken and 'chat' systems), which we have now pursued for over a decade. This includes broad-coverage domain-general natural language processing, dialogue agents built using models of collaborative problem solving, dynamic context-sensitive language modeling, and a rich engineering framework for building dialogue systems in new domains in short times. We can build robust spoken dialogue systems in significantly less time than it would take to collect and annotate a small starter corpora that would be needed for machine-learning driven approaches.



Note, there is a fairly complete set of publications from my research group since 1990.

A Generic Framework for Dialogue Systems

Our research in discourse is focused on two-person extended dialogs in which the speakers have specific tasks to accomplish. An emphasis in this work on developing a theory of dialogue as a collaborative problem solving activity, where the current problem solving situation is used to solve problems in semantic interpretation and the recognition of the intentions underlying the speakers' utterances. Highlights of work in this area include the development of the first computational model of speech acts, the development of a multi-level plan-based analysis involving discourse, and the development of an overall architecture for dialogue systems driven by a collaborative problem solving agent. While it is important for work to be formally well-defined and understood, it is equally important that computational theories can lead to effective implementations. We have demonstrated and tested our models in a wide range of different applications. Most recently, we have been focusing on task/workflow learning systems in which the the system learns a task model from a dialogue with the user that includes a single demonstration of the task. By combining deep language understanding, reasoning, learning and dialog, we can learn robust task models in a matter of minutes. An overview of this work can be found in a paper PLOW that won the outstanding paper award at AAAI in 2007 (videos of PLOW are available here. More details on all our dialogue systems can be found on our projects page.

Selected Publications in Dialogue-based Task Learning

- Allen, J. F., N. Chambers, et al. (2007). PLOW: A collaborative task learning agent. Named Best Paper, National Conference on Artificial Intelligence (AAAI). Vancouver, BC.
- Jung, H., J. F. Allen, et al. (2007). "Utilizing natural language for one-shot task learning." Logic and Computation 18(3): 475-493.
- Chambers, N., J. Allen, et al. (2006). Using Semantics to Identify Web Objects. 21st National Conference on Artificial Intelligence (AAAI-06), Boston, MA.[pdf]

Work on Speech Act-based Dialogue Models

IPR2022-01289 Page 00094

- Allen, J. F. and C. R. Perrault (1980). "Analyzing Intention in Utterances." Artificial Intelligence 15(3).
- Perrault, C. R. and J. F. Allen (1980). "A Plan-based Analysis of Indirect Speech Acts." Computational Linguistics 6(3): 167-182.
- Allen, J.F. and D.J. Litman. "Discourse Processing and Common sense Plans." in P.R. Cohen, J. Morgan, and M. Pollack, Intentions and Communication, MIT Press, 1990.
- Hinkelman, E. and J.F. Allen. ``Two Constraints on Speech Act Ambiguity.'' In Proc., 27th Meeting of the Assoc. for Computational Linguistics, 1989.
- Traum, D. and Allen, J.F. "Discourse Obligations in Dialogue Processing" Proc. 32nd Meeting of the Assoc. for Computational Linguistics, 1994.
- Peter Heeman and James Allen, Speech Repairs, Intonational Phrases and Discourse Markers: Modeling Speakers' Utterances in Spoken Dialog, Computational Linguistics, Vol. 25-4, 1999.

Papers on Selected TRIPS-based Dialogue Systems

- Allen, J.F., B. Miller, E. Ringger and T. Sikorski A Robust System for Natural Spoken Dialogue, Proc. 34th Meeting of the Assoc. for Computational Linguistics, 1996.
- George Ferguson and James F. Allen, TRIPS: An Integrated Intelligent Problem-Solving Assistant, Proceedings of the Fifteenth National Conference on AI (AAAI-98), Madison, WI, 26--30 July, 1998
- Allen, J., Byron, D., Dzikovska, M., Ferguson, G., Galescu, L. and Stent, A., Towards a Generic Dialogue Shell, Natural Language Engineering,6(3) 2000. pp1-16
- James Allen, Donna Byron, Myroslava Dzikovska, George Ferguson, Lucian Galescu, and Amanda Stent, "Towards Conversational Human-Computer Interaction," AI Magazine, 2001.
- Allen, J. F. and G. Ferguson (2002). Human-Machine Collaborative Planning. 3rd Int'l. NASA Workshop on Planning and Scheduling for Space, Houston, TX.
- James Allen, George Ferguson, and Amanda Stent, "An architecture for more realistic conversational systems," in Proceedings of Intelligent User Interfaces 2001 (IUI-01), 1-8, Santa Fe, NM, January 14-17, 2001.
- Allen, J. F., N. Blaylock, et al. (2006). "Chester: Towards a personal medical advisor." Biomedical informatics 39(5): 500-513.
- Allen, J. F., N. Chambers, et al. (2007). PLOW: A collaborative task learning agent. Named Best Paper, National Conference on Artificial Intelligence (AAAI). Vancouver, BC

Broad-coverage Deep Language Understanding

For over a decade we have been developing a broad-coverage, domain general natural language parsing system. While it has been constructed in a application-driven manner, with extensions motivated by particular problems in practice, the grammar and lexicon are all designed to be domain-general. We have only one grammar and lexicon that is used is all our different applications. As a result, when we start a new project, we already have a sophisticated language processing system in place at the start of the project, and have no need to collect and anotate corpora in order to get a high performance system. The parser produces a rich semantic representation of sentences suitable for discourse reasonin such as reference resolution and intention recognition. The parser produces logical forms using a generic ontology (the LF ontology) and we have a set of ontology-mapping tools that allows us to map to the target ontology/representation of each application. Feel free to try out the TRIPS Parser. The basic ideas underlying the grammar and parser are best described in my book Natural Language Understanding, 2nd edition. Specific details on the actual system, its uses and evaluations, are found in the references below.

Selected Publications on Deep Language Understanding

- Allen, J.F. Natural Language Understanding, Benjamin Cummings, 1987, Second Edition, 1994.
- Allen, J.F. ``Natural Language, Knowledge Representation and Logical Form.'' In M. Bates, R. Weischedel (eds.), Challenges in Natural Language Processing, Cambridge University Press, 1993.
- Swift, M. D., J. F. Allen, et al. (2004). Skeletons in the parser: Using a shallow parser to improve deep parsing. COLING '04, Geneva, Switzerland.
- Dzikovska, M., M. Swift, et al. (2005). Generic parsing for multi-domain semantic interpretation. International Workshop on Parsing Technologies (Iwpt05), Vancouver BC.
- Allen, J. F., M. Dzikovska, et al. (2007). Deep linguistic processing for spoken dialogue systems. Workshop on Deep Linguistic Processing, Association for Computational Linguistics. Prague
- Dzikovska, M., J. F. Allen, et al. (2007). "Linking Semantic and Knowledge Representation in a Multi-Domain Dialogue System." Logic and Computation 18(3): 405-430.
- Manshadi, M. H., J. F. Allen, et al. (2008). Towards a Universal Underspecified Semantic Representation. 13th Conf. on Formal Grammar. Hamburg, Germany
- Allen, J., M. Swift, et al. (2008). Deep Semantic Analysis of Text. Symposium on Semantics in Systems for Text

IPR2022-01289 Page 00095

Processing (STEP). Venice, Italy.

Reasoning About Action and Time

The research in plan reasoning draws much of its motivation from the dialog work. In particular, the representation of plans must support a wide range of different forms of reasoning: plan construction (i.e. traditional planning), plan recognition, plan evaluation, and the communication of plans between agents. Much of our work in this area has focused on the representation of time and action, and we have reformulated the planning problem as a problem in temporal reasoning. Within this framework, we have developed a representation of plans that is temporally explicit and supports plan construction, recognition and communication. We are also exploring methods of temporal reasoning that are viable even with large data sets of temporal information.

Selected Publications on Temporal Reasoning

- Allen, J.F. ``Maintaining Knowledge about Temporal Intervals.'' Communications of the ACM 26, 11, 832-843, November 1983.
- Allen, J.F. ``A General Model of Action and Time.'' Artificial Intelligence 23, 2, July 1984.
- Allen, J.F. and P.J. Hayes. ``Moments and Points in an Interval-Based Temporal Logic.'' Computational Intelligence, January 1990.
- Allen, J.F. "Time and time again: The many ways to represent time" Int'l. Jr. of Intelligent Systems 6, 4, 341-356, July 1991. ps
- Allen, J.F. and Ferguson, G. Actions and Events in Interval Temporal Logic, J. Logic and Computation 4, 5, 1994.

Selected Publications on Plan Reasoning

- Kautz, H. and Allen, J.F. "Generalized Plan Recognition" In Proc., Proc., AAAI Nat'l. Conf. on Artificial Intelligence, Philadelphia, PA, 1986
- Allen, J.F. Planning as Temporal Reasoning. In Proc., 2nd Principles of Knowledge Representation and Reasoning, Morgan Kaufmann, 1991.
- Allen, J. and J. A. Koomen (1983). Planning using a Temporal World Model. 8th Int'l. Joint Conference on Artificial Intelligence, Karlsruhe, Germany.
- Allen, J.F. et al. Reasoning About Plans, Morgan Kaufmann, 1991.
- Allen, J., N. Blaylock, et al. (2002). A Problem Solving Model for Collaborative Agents. 1st Int'l. Joint Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS-02), Bologna, Italy, ACM Press.[PDF]
- Blaylock, N. and J. F. Allen (2006). Fast hierarchical goal schema recognition. 21st National Conference on Artificial Intelligence (AAAI-06), Boston.[pdf]

Back to URCS Faculty directory

Back to URCS Home Page

IPR2022-01289 Page 00096

ATTACHMENT

M

IEEE.org IEEE Xplore IEEE SA IEEE Spectrum More Sites SUBSCRIBE SUBSCRIBE Cart Create Account Perso Sign

Browse My Settings Help Institutional Sign In

Institutional Sign In

All

ADVANCED SEARCH

Using Broad Phonetic Group Experts for Improved Speech Recognition

Publisher: IEEE

Cite This PDF

Patricia Scanlon ; Daniel P. W. Ellis ; Richard B. Reilly All Authors

27 Paper Citations

1 Patent Citation

265 Full Text Views

Alerts

Manage Content Alerts

Add to Citation Alerts

Abstract

Document Sections

I. Introduction

II. Mixtures of Experts

III. Mutual Information

IV. Classifier Architecture

V. Classifier Experiments

Show Full Outline

Authors

Figures

References

Citations

Downl PDF

Abstract: In phoneme recognition experiments, it was found that approximately 75% of misclassified frames were assigned labels within the same broad phonetic group (BPG). While the... **View more**

Metadata

Abstract:

In phoneme recognition experiments, it was found that approximately 75% of misclassified frames were assigned labels within the same broad phonetic group (BPG). While the phoneme can be described as the smallest distinguishable unit of speech, phonemes within BPGs contain very similar characteristics and can be easily confused. However, different BPGs, such as vowels and stops, possess very different spectral and temporal characteristics. In order to accommodate the full range of phonemes, acoustic models of speech recognition systems calculate input features from all frequencies over a large temporal context window. A new phoneme classifier is proposed consisting of a modular arrangement of experts, with one expert assigned to each BPG and focused on discriminating between phonemes within that BPG. Due to the different temporal and spectral structure of each BPG, novel feature sets are extracted using mutual information, to select a relevant time-frequency (TF) feature set for each expert. To construct a phone recognition system, the output of each expert is combined with a baseline classifier under the guidance

More Like This

A real-time weighted-eigenvector MUSIC method for time-frequency analysis of electrogastrogram slow wave

2010 Annual International Conference of the IEEE Engineering in Medicine and Biology

Published: 2010

Spectrogram time-frequency analysis and classification of digital modulation signals

2007 IEEE International Conference on Telecommunications and Malaysia International Conference on Communications

Published: 2007

Show More

IEEE websites place cookies on your device to give you the best user experience. By using our websites, you agree to the placement of these cookies. To learn more, read our Privacy Policy.

Accept & Close

Citations
Keywords
Metrics
More Like This

of a separate BPG detector. Considering phoneme recognition experiments using the TIMIT continuous speech corpus, the proposed architecture afforded significant error rate reductions up to 5% relative

Published in: IEEE Transactions on Audio, Speech, and Language Processing (Volume: 15 , Issue: 3, March 2007)

Page(s): 803 - 812 **INSPEC Accession Number:** 9327478

Date of Publication: 20 February 2007 **DOI:** 10.1109/TASL.2006.885907

Publisher: IEEE

ISSN Information:

IEEE Keywords

Speech recognition, Mutual information, Context modeling, Feature extraction, Data mining, Time frequency analysis, Acoustic signal detection, Detectors, Error analysis

INSPEC: Controlled Indexing

error statistics, signal classification, speech recognition, time-frequency analysis

INSPEC: Non-Controlled Indexing

broad phonetic group experts, speech recognition, phoneme recognition, acoustic models, temporal context window, time-frequency feature set, phone recognition system, baseline classifier, TIMIT continuous speech corpus, error rate reductions

Author Keywords

Automatic speech recognition, broad phonetic groups (BPGs), mixture of experts, mutual information (MI)

Contents

I. Introduction

The fundamental task of the acoustic model in a speech recognizer is to estimate the correct subword or phonetic class label for each frame of the acoustic signal. The phoneme can be defined as the smallest phonetic unit in a language that is capable of conveying a distinction in meaning; however, phonemes that may be within the same broad phonetic group (BPG) contain very similar temporal characteristics and can be easily confused. In phoneme recognition experiments on the TIMIT database, reported in [5], it was observed that almost 80% of all misclassified frames are identified as phonemes within the same BPG as the correct target. The BPGs in these experiments were vowels, stops, weak fricatives, strong fricatives, and nasals.

Sign in to Continue Reading

Authors
Figures

IEEE websites place cookies on your device to give you the best user experience. By using our websites, you agree to the placement of these cookies. To learn more, read our Privacy Policy.
Accept & Close

IPR2022-01289 Page 00099

Figures

References

Citations

Keywords

IEEE Keywords

Speech recognition, Mutual information, Context modeling, Feature extraction, Data mining, Time frequency analysis, Acoustic signal detection, Detectors, Error analysis

INSPEC: Controlled Indexing

error statistics, signal classification, speech recognition, time-frequency analysis

INSPEC: Non-Controlled Indexing

broad phonetic group experts, speech recognition, phoneme recognition, acoustic models, temporal context window, time-frequency feature set, phone recognition system, baseline classifier, TIMIT continuous speech corpus, error rate reductions

Author Keywords

Automatic speech recognition, broad phonetic groups (BPGs), mixture of experts, mutual information (MI)

Metrics

IEEE Personal Account

CHANGE USERNAME/PASSWORD

Purchase Details

PAYMENT OPTIONS

VIEW PURCHASED DOCUMENTS

Profile Information

COMMUNICATIONS PREFERENCES

PROFESSION AND EDUCATION

TECHNICAL INTERESTS

Need Help?

US & CANADA: +1 800 678 4333

WORLDWIDE: +1 732 981 0060

CONTACT & SUPPORT

Follow

About IEEE *Xplore* | Contact Us | Help | Accessibility | Terms of Use | Nondiscrimination Policy | IEEE Ethics Reporting | Sitemap | Privacy & Opting Out of Cookies

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

© Copyright 2022 IEEE - All rights reserved.

IEEE Account

» Change Username/Password

Purchase Details

» Payment Options

» View Purchased Documents

Profile Information

» Communications Preferences

» Technical Interests

Need Help?

» **US & Canada:** +1 800 678 4333

» Contact & Support

IEEE websites place cookies on your device to give you the best user experience. By using our websites, you agree to the placement of these cookies. To learn more, read our Privacy Policy.

Accept & Close

About IEEE *Xplore* | Contact Us | Help | Accessibility | Terms of Use | Nondiscrimination Policy | Sitemap | Privacy & Opting Out of Cookies

IPR2022-01289 Page 00100

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.
© Copyright 2022 IEEE - All rights reserved. Use of this web site signifies your agreement to the terms and conditions.

IEEE websites place cookies on your device to give you the best user experience. By using our websites, you agree to the placement of these cookies. To learn more, read our Privacy Policy.

Accept & Close

ATTACHMENT

N

Public Catalog

Copyright Catalog (1978 to present)
Search Request: Left Anchored Title = IEEE Transactions on Audio, Speech and Language Processing
Search Results: Displaying 3 of 39 entries

previous next

Labeled View

IEEE Transactions on audio, speech, and language processing.

Type of Work: Serial
Title: IEEE Transactions on audio, speech, and language processing.
Serial Publication Year: 2007
ISSN: 1558-7916
1063-6676
Description: print material.
Frequency: Pub. 6 times per year.
Continues: IEEE Transactions on speech and audio processing = ISSN 1063-6676.
Description based on: Vol. 14, no. 1, Jan. 2006.
Copyright Claimant: IEEE.
Issues Registered: v. 15, no. 2, Feb07. Pub. 2007-01-17; Reg. 2007-06-12; TX0006610997
v. 15, no. 2, Feb07. Pub. 2007-01-17; Reg. 2007-06-22; TX0006612701
v. 15, no. 3, Mar07. Pub. 2007-02-14; Reg. 2007-06-12; TX0006610997
v. 15, no. 3, Mar07. Pub. 2007-02-14; Reg. 2007-06-22; TX0006612701
v. 15, no. 6, Aug07. Pub. 2007-07-20; Reg. 2007-11-28; TX0006643713
v. 15, no. 6, Aug07. Pub. 2007-07-20; Reg. 2008-02-14; TX0006663644
v. 15, no. 7, Sep07. Pub. 2007-08-14; Reg. 2008-02-14; TX0006663644
v. 15, no. 7, Sep07. Pub. 2007-08-14; Reg. 2007-11-28; TX0006643713
v. 15, no. 8, Nov07. Pub. 2007-10-12; Reg. 2007-11-28; TX0006643713
v. 15, no. 8, Nov07. Pub. 2007-10-12; Reg. 2008-02-14; TX0006663645
v. 16, no. 1, Jan08. Pub. 2007-12-19; Reg. 2008-02-14; TX0006663645
Other Title: Transactions on audio, speech, and language processing
Names: IEEE
Institute of Electrical and Electronics Engineers

previous next

Save, Print and Email (Help Page)

Select Download Format Full Record Format for Print/Save

Enter your email address: Email

Help Search History Titles Start Over

Contact Us | Request Copies | Get a Search Estimate | Frequently Asked Questions (FAQs) about Copyright | Copyright Office Home Page | Library of Congress Home Page

ATTACHMENT

O



Speech Perception Using Real-Time Phoneme Detection: The BeBe System

by Latanya Sweeney and Patrick Thompson

Abstract

We define a new approach to speech recognition based on auditory perception and modeled after the human brainÕs tendency to automatically categorize speech sounds [House 1962; Liberman 1957]. As background, todayÕs speech recognition systems are knowledge-driven since they require the existence of word and syntax-level knowledge to identify a word from the sound. In contrast, our system uses no higher-level knowledge. Its architecture consists of competing parallel detectors which in real time identify phonemes in the waveform. Each detector, which is a simple algorithm, continuously samples the sound and reports the degree to which the samples contain its designated phoneme. The phoneme detector with the highest precedence and the greatest certainty above a minimal threshold prevails and its phoneme is added to an output queue. In preliminary experiments, four such detectors were tested and they properly identified 83-100% of their designated phonemes in both discrete and continuous speech, independent of the speaker, suggesting that an overall system which incorporates our approach would be much more robust and flexible than traditional systems.

Citation:
Latanya Sweeney and Patrick Thompson. Speech Perception Using Real-Time Phoneme Detection: The BeBe System. Massachusetts Institute of Technology, Laboratory for Computer Science: Tech Report MIT-LCS-TR-736. 1998. (PDF).

Related Links

- Other Projects at the Data Privacy Lab
- Data Privacy Laboratory

Copyright © 2011. President and Fellows Harvard University. | IQSS | Data Privacy Lab | [info@dataprivacylab.org]

ATTACHMENT

P

Speech Perception Using Real-Time Phoneme Detection: The BeBe System

Latanya Sweeney and Patrick Thompson

Laboratory for Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
sweeney@lcs.mit.edu, pmt@ai.mit.edu

© Copyright 1997 by Latanya Sweeney and Patrick Thompson

Speech Perception Using Real-Time Phoneme Detection: The BeBe System

Latanya Sweeney and Patrick Thompson

Laboratory for Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
sweeney@lcs.mit.edu, pmt@ai.mit.edu

Abstract

We define a new approach to speech recognition based on auditory perception and modeled after the human brain's tendency to automatically categorize speech sounds [House 1962; Liberman 1957]. As background, today's speech recognition systems are knowledge-driven since they require the existence of word and syntax-level knowledge to identify a word from the sound. In contrast, our system uses no higher-level knowledge. Its architecture consists of competing parallel detectors which in real time identify phonemes in the waveform. Each detector, which is a simple algorithm, continuously samples the sound and reports the degree to which the samples contain its designated phoneme. The phoneme detector with the highest precedence and the greatest certainty above a minimal threshold prevails and its phoneme is added to an output queue. In preliminary experiments, four such detectors were tested and they properly identified 83-100% of their designated phonemes in both discrete and continuous speech, independent of the speaker, suggesting that an overall system which incorporates our approach would be much more robust and flexible than traditional systems.

Keywords: speech recognition, phoneme detection, speech enhancement, auditory perception, AI architectures, speech understanding, real-time systems.

Acknowledgments

We gratefully acknowledge Gerald Sussman for discussions, support and enthusiasm. We also thank Norman Margolus, Ken Yip, Ken Stevens, Robert Berwick, Gina Levow, Tom Knight and Patrick Winston for discussions; Eric Jordan and Philipp Schmid for editorial comments; and Peter Szolovits for providing an environment that made it possible for us to explore our own ideas. Henry Leitner inspired this work and we are grateful for his continued support and for that of Harvard University DCE. This work has been further supported in part by a Medical Informatics Training Grant (1 T15 LM07092) from the National Library of Medicine.

Contents

1 Introduction **1**

1.1 The BeBe System 1

2 Background **2**

2.1 Computational architectures 2

2.2 Continuous speech 4

2.3 Phoneme models 6

2.4 Human speech recognition 6

2.5 Students learn phonemes 7

3 Design **8**

4 Implementation **9**

5 Results **12**

5.1 Comparison of others 14

6 Discussion **15**

6.1 BeBe as the first stage 16

6.2 Recognizing vowels 17

6.3 Assessment and future work 18

References **19**

IPR2022-01289 Page 00112

1 Introduction

The task of speech recognition is to map a digitally encoded signal to a string of words. Over the past 10 years speech recognition technology has advanced dramatically, evolving into 65,000-word vocabulary research systems capable of transcribing naturally spoken sentences on specific topics from any new talker [Bourlard, et al., 1996], but achieving human-like performance remains distant. Lippmann [1996] points out: (1) the error rates of machines are more than an order of magnitude greater than those for humans under the most ideal circumstances for the machines; (2) machine performance plummets much faster than humans when operating in noise and other degraded conditions; (3) humans exhibit much more powerful types of adaptation and incorporate newly learned words; and (4) humans rely on context much less than machines and can accurately recognize nonsense words, which are words that sound like English words but in reality have no meaning.

We agree with Lippmann that improvements at the low, acoustic-phonetic level must be achieved if machines are to equal human performance on real-world tasks. Our challenge is to model human auditory reasoning and behavior, so we consider our approach in this work as one of auditory perception, which models how the identities of sounds, including speech sounds, are learned and processed by a human listener.

1.1 The BeBe System

The specific aim of the BeBe System is to reliably and consistently detect phonemes in continuous speech. Computation can be performed in real time and recognition is robust enough to adapt to different speakers and changes in talking speed. Phonemes are the basic sound groups of a language and most languages have 50 or fewer phonemes, so explicitly identifying phonemes in the waveform offers an economic representation.

Phonemes are language-specific, so the identification of phonemes in languages other than American English requires building different phoneme detectors. Also, there is tremendous variation among sounds within a phoneme, so we consider phonemes to be abstract classifications and the goal of the BeBe System to be the continuous classification of waveforms. BeBe's activity concerns what is traditionally termed the *preprocessing stage* of speech recognition, in which the original sound waveform is converted to a phonemic digital representation. Depending on the speech recognition system, there may be three or more stages

IPR2022-01289 Page 00114

until a word or phrase is recognized where each stage often inherits inexact and noisy results from earlier stages. As a result, ambiguity in the overall system can magnify exponentially. Having a reliable phoneme-based detector in the preprocessing stage clearly improves speech recognition systems since it dramatically reduces uncertainty throughout the system and in its best case reduces the speech recognition problem to one of looking up phoneme sequences in a table.

2 Background

One of the issues impeding widespread implementation of large-vocabulary, continuous-speech systems is computational complexity [Rabiner and Juang, 1993]; therefore, in providing background for our approach, we will first discuss computational architectures. Then we will look at continuous speech and the acoustic-phonetic approach to recognition since research in these areas also views words as sequences of phonemes. Following that, we will shift our attention to human speech perception.

2.1 Computational architectures

In the 1970s, the Advanced Research Projects Agency (ARPA) of the United States Department of Defense conducted its Speech Understanding Research Project (ARPA SUR) which catapulted speech recognition research from speaker-dependent, small-wordlist recognition to the large-scale language model systems available today. Only one system met ARPA SUR's original mandate, Carnegie Mellon University's Harpy [Klatt, 1977], and the success of Harpy's statistical modeling techniques continues to have a profound effect as researchers seek to build larger statistical knowledge bases in an attempt to overcome problems and extend the performance of systems. Some attention has been given to recognizing phonemes directly from the waveform in the preprocessing stage, but little attention has focused on the use of parallelism in the preprocessing stage to improve speed and accuracy and to mitigate the overall system's computational complexity.

Our BeBe System utilizes numerous detection algorithms competing in parallel to label contiguous samples of sound as being particular phonemes. Each detection algorithm recognizes a specific phoneme though there may be more than one detection algorithm for a phoneme. For example, there is a single detection algorithm for the phoneme /R/ which is found in words like

IPR2022-01289 Page 00115

early, hurt and stir. There is a detection algorithm for the sound [p] as in bumper, spit and culprit, and another detection algorithm for [p^h] as in pit, pain and part, since these two allophones are different sounds of the same phoneme /p/. Our convention is to write phonemes using Arpabet letters (ARPA's phonetic alphabet that can be typed using a traditional computer keyboard) between slashes (/ /) and to write phonetic symbols between square brackets ([ ]).

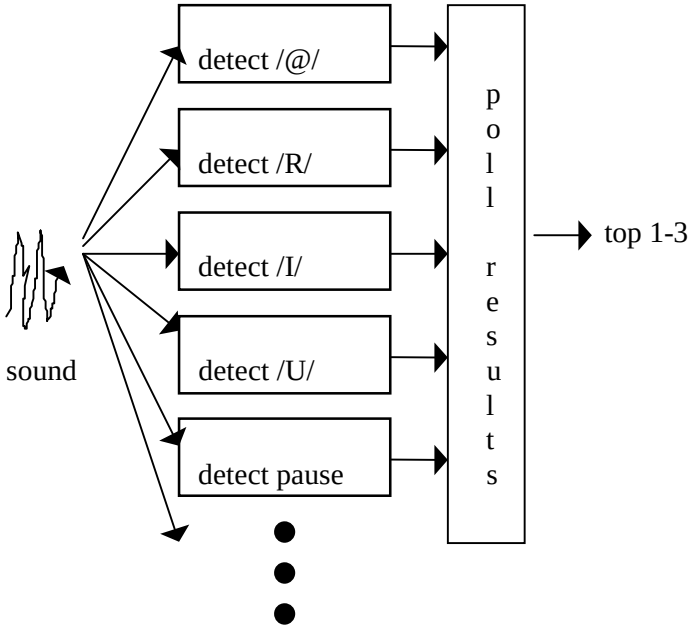


Diagram 1. Block diagram of BeBe System. The detection algorithms report the likelihood that the sound samples contain the designated phoneme. Examples of the sounds are: /@/ as in bat, /R/ as in stir, /I/ as in fit, and /U/ as in should.

Detection algorithms in BeBe share results and compete based on the certainty of their findings. Each algorithm tries to identify occurrences of its assigned phoneme, and reports for each sample in the waveform how likely it is that the sample is part of an instance of the algorithm's assigned phoneme. The algorithm with the highest precedence and the greatest certainty above a minimal threshold prevails and its results may be made available to all detection algorithms for future use. Diagram 1 presents an overview of the BeBe architecture which will be discussed in further detail in the Design and Implementation sections.

In the area of speech recognition, Hearsay-II's blackboard architecture [Erman, et al., 1980] engages multiple knowledge sources that work in parallel. Adjacent sources communicate with each other using a message center called a blackboard. This is similar to the BeBe System except

IPR2022-01289 Page 00116

communication is central to Hearsay-II because each level is believed to be so uncertain that a collaborative effort is required and not a competitive one.

Another approach similar to BeBe's is the Scrub System [Sweeney, 1996] which locates and replaces personally identifying information in medical records. Letters between physicians and notes written by clinicians often contain nicknames, phone numbers and references to other care-takers and family members, making it difficult to share medical records while still maintaining a commitment to patient confidentiality. The Scrub System used numerous algorithms competing in parallel to identify personal information in unrestricted text, and the system found 99-100% of these references. In contrast, the straightforward approach of global search-and-replace properly located no more than 30-60% of all such references. Although the Scrub System used numerous knowledge sources such as lists of area codes, first names, medical terms and so forth, BeBe has no stored lists and uses no higher-level predictive knowledge. BeBe relies only on its ability to recognize phonemes.

In Ether [Kornfeld, 1979], decentralized parallel processing was shown to be an effective alternative, in computational complexity terms, to many kinds of heuristic search strategies that implemented backtracking. Parallelism in Ether, as in BeBe, is design-based and does not necessarily require parallelism in its implementation. However, Ether does not use certainty factors as a scoring system between competing processes; instead the first process to complete the task solves the problem.

2.2 Continuous speech

In continuous speech, the speaker communicates in a natural manner with naturally occurring pauses. Most of the speech recognition systems commercially available today are really "connected speech" systems which require a deliberate pause between each word [Markowitz, 1996]. The pause cannot be eliminated since it is used to identify word boundaries. Table 1 shows a phonetic transcript of some American English phrases and their corresponding text with and without word boundaries. To combat this problem in continuous speech, many systems use triphones, a phoneme surrounded by contextual information on both sides, to model cross-word coarticulation. Researchers at AT&T Bell Laboratories [Pieraccini, et al., 1991] reported that when these highly detailed speech units were used the complexity of the overall implementation increased quadratically with the number of units, making a full-search implementation at that time "totally impractical, if not impossible."

IPR2022-01289 Page 00117

Clearly, having a reliable phoneme detector changes the nature of speech recognition and reduces computational complexity. If a phoneme detector could provide a phonetic transcript that was 100% accurate in connected speech, then the remaining task for speech recognition would be to simply look up the phoneme sequence in a table which, like desktop dictionaries, associates phonetic pronunciation to spellings. As the phoneme detector becomes less reliable, however, the complexity of the remaining task increases. At, say, 90% accuracy, the task may be simple; at 80% accuracy, it may be reasonable; but, at 60% accuracy, recognition from the transcript would be difficult, depending of course on the nature of the inaccuracy. A similar computational complexity relationship holds for continuous speech as well. So, an approach to speech recognition we propose is to build a reliable phoneme detector.

	Phrase I
Phonetic Transcript	[gɛtθIkbʊk]
English text stream	getthickbook
English sentence	get thick book

	Phrase II
Phonetic Transcript	[sɔp^howzpɔrejdkɔrowd]
English text stream	supposeparadecorrode
English sentence	suppose parade corrode

	Phrase III
Phonetic Transcript	[sp^howzpr_1ejdkr_1owd]
English text stream	supposeparadecorrode supposeprayedcrowed
English sentence	suppose prayed crowed, suppose parade corrode

Table 1. A phonetic transcript of English phrases and their corresponding text with and without word boundaries. The last two phrases are the same utterances except phrase II is at normal speed and III is spoken rapidly.

BeBe's processing is a left-to-right, one-pass system. The entire waveform does not need to be saved. As a result, many more word boundaries are explicitly detected than is possible using traditional recognizers. In fact, one of the detection algorithms in BeBe is itself a pause detector reporting how likely it is that the sound sample is a pause. Of course the pause detector in BeBe still will not distinguish all word boundaries. These word divisions must be determined from the phonetic transcript BeBe produces. In the discussion section, we will present strategies that combat this, but first we will examine other phoneme-centered systems, human speech perception, and BeBe's implementation and run-time results.

IPR2022-01289 Page 00118

2.3 Phoneme models

Acoustic-phonetic recognition involves global reasoning about the identity of phonemes in a digitized representation of a spectrogram [Cole, et al., 1980; Zue, 1985]. There are three stages: feature extraction, segmentation and labeling, and word-level recognition. First, the system examines the signal representation for features that describe spectral patterns. Extracted features are then interpreted using acoustic-phonetic rules that attempt to label the phoneme and segment where the phoneme begins and ends, but this is often very uncertain since the feature set does not distinguish well enough between similarities in phonemes and coarticulation effects to be reliable. The uncertainty in the results leads to a set of hypotheses that are organized into a decision tree and the system then searches through its vocabulary for words that match the hypotheses. In the next sections, we will consider whether humans employ this kind of higher-level reasoning to recognize basic sounds.

2.4 Human speech recognition

The brain tends to automatically categorize speech sounds. Experiments conducted by House [1962] showed that when listeners hear nonspeech synthetic sounds which are gradually made more speechlike, an abrupt boundary is found, where on one side the sounds are perceived as speech and on the other side they are not. Further, in experiments by Liberman [1957] a phoneme in synthesized words was gradually varied until it became another phoneme, for example “bad” to “gad” and “pit” to “bit.” Listeners did not hear this gradual variation, but instead made sharp

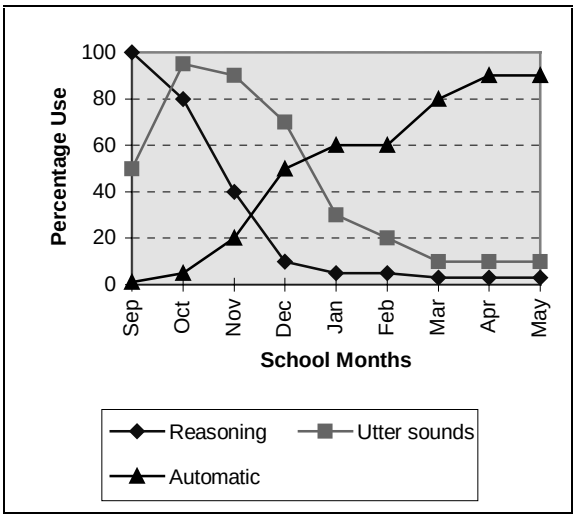


Table 2. Averaged results from retrospective survey. Subjects were asked to plot, for each month, the percentage of time they employed each modality in learning the sounds of a second language.

distinctions among initial consonants. It appears that there is an innate predisposition in the human nervous system to immediately classify speech inputs. Certainly initial training is required to develop these explicit categories, but once present, sound recognition seems an automatic process.

2.5 Students learn phonemes

We recently interviewed and surveyed students learning a foreign language that had a significantly different sound system from their native language. In particular, these were American English speakers learning the Korean language over two consecutive school semesters. The students began the school year with only knowledge of English.

Two students were interviewed separately. The students: (1) described a time during which they had to train their ear to recognize the sounds of the new language, claiming that at first they engaged a lot of reasoning and predictive knowledge to identify sounds (which we refer to as "reasoning"); (2) claimed they uttered competing words in their minds to compare different sounds (which we refer to as "uttering"); and (3) agreed that once the sounds became familiar to their ear, detection appeared to become automatic even if the meanings of the words were unknown (which we refer to as "automatic").

We then surveyed four students and asked them to retrospectively plot how much they used: reasoning, uttering, and automatic detection over the school year. The students plotted values for each month, for each of the three curves. The averaged results are shown in Table 2 and can be interpreted as follows. When the first term began, reasoning was relied on since students could not distinguish Korean sounds at all; context and higher-level knowledge were critical. Several Korean phonemes sound similar, such as /m/ and /n/ in English, so the students often uttered competing sounds internally. While speech perception relies on the listener's familiarity with the language, the talker's characteristics and more, the recognition of speech sounds appeared automatic once the sound categories were established.

In the previous survey we can draw analogies to all three approaches in speech recognition. Acoustic-phonetic recognition, where one reasons about the sound, is similar to the deliberation the students undertook when contemplating what sounds they heard. Likewise, the stochastic approach, as was spawned by Harpy, where one makes comparisons to known sounds and has expectations of what sound is next based on known words, is similar to the uttering practice described by the students who compared similar sounds in their heads. Our goal with the BeBe System is to model the human ability to directly categorize sound in real time without higher-

order reasoning. The students described employing this behavior once they had acquired the ability to categorize the speech sounds of the new language. The recognition of the sounds became automatic.

Even though these survey results may not be statistically or psychologically persuasive, they do agree with the findings of Fletcher and his colleagues, who studied the principles behind human speech recognition from 1918 to 1950 at Bell Labs [Allen, 1994] and concluded that humans decode speech sounds into independent units at an early stage, before semantic context is used.

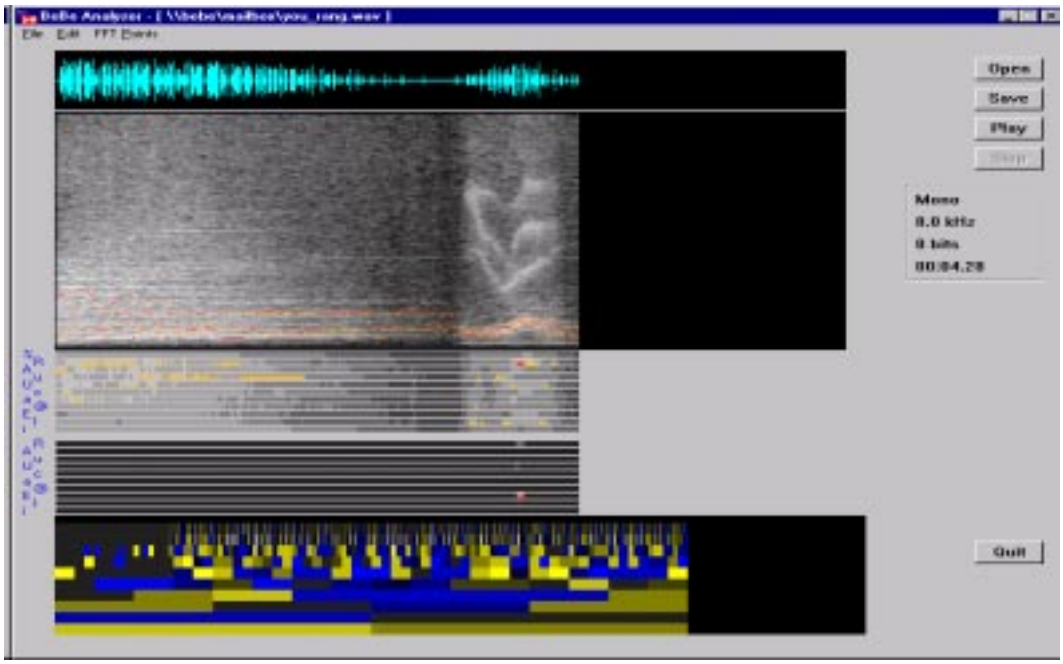


Diagram 2. A presentation of $BeBe_0$ at work. There are five display areas, from top to bottom: the waveform, the FFT spectrogram, the results from the detection algorithms, other information from the detection algorithms, and a wavelet transform.

3 Design

We sought to model the human ability to automatically categorize speech sounds since this did not require a semantic model of the language. Instead it involved an almost "biological" detection of basic sounds directly from the waveform. Further, we wished to evaluate the Scrub architecture when applied to the well-suited problem of phoneme detection.

For each phoneme, or basic sound category, there are one or more detection algorithms, each with precedence based on the number of samples that constitute the algorithm's assigned sound; detectors for phonemes with longer duration have higher precedence. For each sample in the

IPR2022-01289 Page 00121

sound the detection algorithm with the highest precedence reporting the greatest likelihood above a threshold value is considered to have identified an instance of its phoneme. In development, detection algorithms can be executed sequentially in order of precedence to avoid parallel execution. Diagram 1, presented earlier, provides an overview. All detectors report their findings to the polling algorithm, which then makes final decisions based on the certainty of the detectors, their precedence, and time analysis of past results.

Often a separate detection algorithm is used for word-final than for word-initial occurrences of consonants in words since systematic relationships exist between phonemes in particular positions within a word and their pronunciation. Consider the [p] and [p^h] allophones of /p/ presented earlier. Some detectors classify information about background noise. These detectors continuously report findings. There are also special detectors like those that determine whether human speech is present and having these results reduces the number of false positives. At run-time the user can set the threshold and use of these special detectors.

Knowing what instances have already been found in the sound can be useful in reducing ambiguity. For example, if the system encounters the phoneme sequence < /i/ /I/ /i/ > in contiguous sound samples, where the duration of each phoneme is less than its typical length, then the /i/ result can prevail with more certainty and account for all these samples. This type of drift in the speaker's voice is common. Similar sequence inconsistencies occur due to coarticulation at word and syllable boundaries. Thus, maintaining a queue of recent detector results allows the polling algorithm to summarize a series of findings as being a single instance of one phoneme.

The queue used by the polling algorithm can grow to hold a one-second history or about 100 detector results. In producing a phonetic transcript, the polling algorithm outputs one instance of a phoneme for roughly 20 or more contiguous results. Detectors primarily base their decisions on frequency information while the polling algorithm resolves further uncertainty using time analysis.

4 Implementation

We implemented a version of the BeBe System, termed $BeBe_0$, in C++ on a Pentium Pro processor machine running at 200 MHz with 32MB RAM and executing the Windows 95 operating system. A Shure SM10A microphone was used in a SoundBlaster-equivalent audio card with software adapted from Rimmer [1995]. The recording environment was an office setting in which background noise included a loud ventilation system and the electronic hum of

IPR2022-01289 Page 00122

three computer systems. The sound pressure level was estimated at 63 dBA where a typical office with a single computer is around 45-50 dBA [Rabiner, et al., 1993].

Here is a walk through the system. A forward-sliding window identifies 512-sample segments of the audio waveform, nominally sampled at 22050 Hz, monaurally, with 8-bit resolution. Spectral slices are developed by passing these segments through a Hamming window to an FFT analyzer. The moduli of the transformed values are approximately weighted according to the frequency response of the human ear [Jones, 1993] and smoothed along the time axis by a simple recursive filter.

With these steps completed, each detection algorithm has available to it a vector of the sample segment, and vectors of the FFT before and after weighting. The average spectral density for the weighted FFT is also computed (by averaging the values in the vector up to the Nyquist frequency) and posted.

Each detector is free to use any of this information to determine its certainty that the sample is an instance of its phoneme. The certainty factor can be conceived as a number between 0 and 1. For many of the vowels, we found that simple heuristics concerning typical duration and templated spectral-band densities for formants F1, F2, and F3 [Peterson and Barney, 1952] could accurately identify instances. False positives were gated out by requiring a minimum average spectral density (a loudness threshold) and then a minimum ratio of template-band density to average density for suspected vowels (a minimum vowel-signal-to-noise). Algorithm 1 provides a

Given a weighted, 512-segment FFT *wfft*, its average spectral density d_{avg}, and sampling frequency 22050 Hz, the response of the /R/ detector is determined as follows:

a. The computation of the band spectral density d_{band} is based on the averaged formants for /R/ from Peterson and Barney [1952]. The index i of a frequency F in vector *wfft* is determined by:

$$i = \frac{512}{22050} F .$$

Then,

$$d_{\text{band}} = \underbrace{\sum_{i=10}^{11} wfft[i]}_{\text{F1}\in[490\pm23]} + \underbrace{\sum_{i=30}^{32} wfft[i]}_{\text{F2}\in[1350\pm29]} + \underbrace{\sum_{i=38}^{39} wfft[i]}_{\text{F3}\in[1690\pm29]} .$$

b. If $d_{\text{band}} < 50$, then exit and return 0 since the minimal speech signal is not present.

c. Let $b = (d_{\text{band}} - 3d_{\text{avg}})$. If $b \le 0$ then return 0; else, let $b = \frac{1}{6} b d_{\text{avg}}$. If $b > 1$ then let b=1. Return b as the certainty factor.

Algorithm 1. The non-adaptive version of the /R/ detector as used in BeBe_0. An example of the /R/ sound is found in st<u>ir</u>. The magic values 50, 3 and 6 were determined empirically.

IPR2022-01289 Page 00123

listing of the /R/ detector that uses this simple strategy. Other detectors work differently, and some could be made adaptive.

$BeBe_0$ detects formants based on fixed templates. A logarithmically varying template based on critical bands [Schroeder, et al., 1979], such as the Bark scale used by Huang [1991], would likely improve detector accuracy and speaker independence in BeBe.

$BeBe_0$ gets robust results with only the fixed-template version of the /R/ detector. If one were to build a BeBe System with only /R/ and /U/ detectors, where the /U/ detector is similarly constructed, the number of correct detections for /R/ when present would remain high, but the number of false detections, where /R/ is detected but not present, would also be very high. Consider the definitions for correctness and accuracy listed in Definition 1. In the BeBe architecture, having detectors that are somewhat exclusive and correct, even if they are not accurate, appears to boost the performance of the overall system.

Diagram 2 shows $BeBe_0$'s detectors at work. From top to bottom are displays of the voltage waveform, the FFT-based spectrogram, two ribbon-sets of results, and a wavelet transform. The sound in the waveform is a loud cymbal crash and then the words "you rang." Of particular note is the upper ribbon-set display, showing the vowel detection for each Arpabet symbol on the left-hand side. The monochrome intensity in each ribbon shows that detector's certainty that it has identified its vowel. During the cymbal crash, the false detections are overruled, since the average spectral density exceeds the band densities of any of the vowels. On a color display, such determinations show as orange ribbons. Later in the waveform, during the words "you rang," the selectivity of the detectors is more apparent; for example, the /R/ line is bright, denoting the presence of the /R/ phoneme. A red ribbon indicates which detector has the highest certainty at points along the waveform. The second ribbon-set display shows a competing group of detectors, with their heuristics still under development, and some of these detectors use information from the wavelet transform.

$$\text{correctness} = \frac{\text{number correct}}{\text{total occurences}}$$

$$\text{accuracy} = \frac{\text{number correct - false positives}}{\text{total occurences}}$$

Definition 1. The BeBe architecture favors independent detectors that are correct but not necessarily accurate.

IPR2022-01289 Page 00124

5 Results

We conducted an experiment to determine how well $BeBe_0$ would identify phonemes in discrete and continuous speech. The subjects were 4 male adults, all native American English speakers, each from a different state: Alabama, Illinois, California and Washington. Each subject was given a page containing 5 lines of text. The first 2 lines had isolated words and the subjects were instructed to pause between each word. The last 3 lines were full sentences and the subjects were to read them aloud in a natural manner. Table 3 contains a copy of these words. One subject did not speak 1 word and another did not speak 2 of the words in the text, so these were not counted.

Bob. Hat. Heard. Think. Thought. Fee.
Head. Book. Boot. Mud. Through.
Kim needs that book first.
She saw the fat blue bird.
His red boot got muddy.

Table 3. The lines above were spoken into $BeBe_0$ by the subjects. The first 2 lines have a pause between each word and the last 3 are continuous sentences.

$BeBe_0$ was tested on its recognition of 4 phonemes: /@/ as in b<u>a</u>t, /R/ as in st<u>ir</u>, /I/ as in f<u>i</u>t, and /U/ as in sh<u>ou</u>ld. These phonemes were chosen because they cover each vocal-tract articulatory configuration for vowels: /I/ and /@/ are considered front vowels, /R/ is a middle vowel and /U/ is a back vowel.

	R	A	u	U	c	a	@	E	I	i
/R/	**12**									
/A/		5				1				
/u/	1		9	1						
/U/		**1**		**7**						
/c/				1	8				1	
/a/						6	1			
/@/					**1**	**1**	**11**			
/E/	1							7	1	
/I/	**1**								**10**	**1**
/i/			1						1	10

Table 4. One review of results from all 10 detectors in $BeBe_0$. Each row denotes a phoneme that occurred in the sound, and each column displays the results from a detector. If the results were perfect, all numbers would appear in the diagonal. The shaded rows highlight the phonemes under study, and is also where both reviews agree.

A total of 10 competing phoneme detectors were used, and these were based on the averaged formants from Peterson and Barney. Tables 4 through 8 have summary results. No training was performed. During the development of $BeBe_0$, we recorded approximately 50 three-second recordings of various sounds and speakers. These files were useful during debugging, but none of

IPR2022-01289 Page 00125

these recordings included the test lines shown in Table 3. Samples for one subject were included, though the subject did not speak the text in Table 3.

The results from all 10 detectors were scored in 2 different reviews, which both agreed on the correctness of the 4 phoneme detectors shown in Table 7, but differed with respect to the correctness of the other detectors. The source of the problem was in transcribing the sounds from the speakers. For example, some subjects pronounced *the* with an /i/ sound while others used an /E/ sound. Table 4 shows a summary of one review. The other review differs in about 7 instances.

Considering the 4 phonemes under study, $BeBe_0$ performed quite well, from 83-100% correct. There was no appreciable difference between isolated and continuous speech. There were a total of 25 phonemes in the text and 4 speakers, totaling 100 possible detections. The detectors reported false positives as follows: /@/ had 1, /R/ and /I/ each had 3 and /U/ had 2, which is 1-3%.

Phoneme	Total Occurrences	Correctness
/R/	**12**	**1.00**
/A/	6	0.83
/u/	11	0.82
/U/	**8**	**0.88**
/c/	10	0.80
/a/	7	0.86
/@/	**13**	**0.85**
/E/	9	0.78
/I/	**12**	**0.83**
/i/	12	0.83

Table 5. Correctness results for data in Table 4. The average correctness overall is 85%.

Detector	Totals	False Positives	Accuracy
R	**15**	**3**	**0.75**
A	6	1	0.67
u	10	1	0.73
U	**9**	**2**	**0.63**
c	9	1	0.70
a	8	2	0.57
@	**12**	**1**	**0.77**
E	7	0	0.78
I	**13**	**3**	**0.58**
I	11	1	0.75

Table 6. Accuracy results for data in Table 4. The average accuracy overall is 69.5%.

	#correct	Total	%correct
/@/	11	13	85
/R/	12	12	100
/I/	10	12	83
/U/	7	8	88

Table 7. Results from human subjects speaking the text in Table 3 with $BeBe_0$ detecting these phonemes. For a given phoneme, %correct = (number identified correctly) / (total occurrences) $\times$ 100.

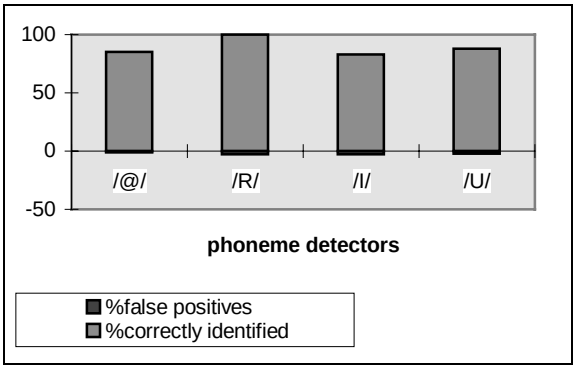


Table 8. Data from Table 7 with false positives.

5.1 Comparison of others

As demonstrated in the previous subsection, evaluating the performance of a speech-based system is quite difficult and comparing different systems is even worse. Different systems are meant to accomplish different goals, and by design perform better on some tests than others. Also, we cannot compare the results of the simple experiment discussed in the previous subsection, whose purpose was to illustrate the plausibility of the BeBe approach, with experiments that report results based on a large corpus over many speakers. Nevertheless, looking at the reported accuracy of other systems helps evaluate the motivation for the BeBe System.

Most of the early experiments focused on vowel recognition. The summary of results listed in Table 9 implies that recognizing vowels is at least as difficult as identifying consonants. Schmid [1996] reports 62% accuracy using a unigram language model and detecting phonemes /t/ and /m/ using a vowel classifier based on the N-best consistent interpretations of formant information.

IPR2022-01289 Page 00127

The results in Table 9 for vowel accuracy are based on the TIMIT corpus, which contains speech from 630 speakers from 8 major dialects of American English, each speaking 10 sentences. As future work, $BeBe_0$ and its adaptive version should be tested on TIMIT sound samples to allow direct comparison.

Type	Researcher	Description	Accuracy
vowel	Meng, et al. [1991]	auditory model	64.5%
vowel	Carlson, et al. [1992]	neural net, gender info	65.6%
phoneme	Chigier, et al. [1992]	neural net	78.0%
phoneme	Digalakis [1993]	male only	73.9%
unigram	Schmid [1996]	n-formants, detect /t/ and /m/ phonemes	61.1%
unigram	Schmid [1996]	n-formants with cepstral	62.0%

Table 9. Accuracy results from various researchers. Vowel findings are based on the TIMIT corpus. The unigram language involves testing with isolated phoneme sounds.

Perceptual experiments using the TIMIT database were conducted by Cole and Muthusamy [1992]. Vowel identification for the 16 TIMIT vowels when isolated from sentences and then played to human subjects had a correctness of 0.55, which rose to 0.66 when acoustic context was provided. Machine recognition of vowels already exceeds this correctness level, which suggests that training of the human auditory system is sensitive to statistical correlation of related local context beyond what was tested. Some consonant-vowel and vowel-consonant transitions, for example, are common and others do not occur at all and such may be related to how humans pronounce and recognize nonsense words [Sweeney, 1996b].

6 Discussion

We have presented the BeBe architecture and demonstrated its robustness. There are about 40 phonemes in English and we expect about 100 detection algorithms in BeBe; though many of them are not yet active, the feasibility of this approach has been illustrated even though our experiment did not utilize the BeBe architecture's ability to employ different algorithms for different phonemes. In closing, we will discuss how BeBe might be used in a speech system, examine other phoneme recognizers, and assess BeBe's performance and potential.

IPR2022-01289 Page 00128

6.1 BeBe as the first stage

Earlier we discussed the problem of word boundaries where the identification of a word even from a proper phonetic transcription requires knowledge of local context at a higher level. We could easily imagine a greedy algorithm that starts processing phonetic symbols matching the best guess at the word so far; if that does not lead to identification of the word and the second word also, then backtrack and take the next best guess at the word. Or, we could employ a statistical or HMM approach to the results from BeBe and the resulting system would be far simpler than traditional HMM models.

Another strategy applies the BeBe architecture at a higher level. A phonetic transcript is sent to a second stage that converts the transcript to poorly spelled English words using sound-to-spelling rules, phoneme-to-grapheme mappings [Hall, 1961], or orthographic rules [Sweeney, 1996]; see Diagram 3. The candidate words from the second stage then go to a spelling corrector which produces the final result. The entries at all stages include likelihood measures, and the most likely entries (no more than seven) are passed to the next stage. The most attractive aspects to these approaches, not including the spelling corrector, are: no large vocabulary is needed for storage and retrieval, and the number of words recognizable is infinite and includes nonsense words such as “throck” and “zat” as well as proper names and terms specific to a profession or field with no additional training. The recognizer, without the spelling corrector, can spell any word that is pronounceably consistent to the language’s pronunciation rules, and with the spelling corrector this can be limited to only the list of known spellings. Also, these approaches adhere to well-defined abstraction layers that have no global methods, and all three approaches to the middle stage could be applied in parallel.

As part of their study of human speech recognition, Fletcher and his colleagues proposed a model of human speech recognition that consisted of a cascade of recognition layers, starting with the cochlea [Allen, 1994]. As in the BeBe architecture, no feedback is assumed between layers; Fletcher’s abstractions are similar. The first layer, based on the cochlea, determines the signal-to-noise ratio in about 2800 overlapping critical band channels; the second layer extracts about 20 speech features from the channel information in a local manner; the next layer maps those features onto phonemes, and then final layers determine syllables and words. In the next sections, we discuss other phoneme recognizers and then assess BeBe’s performance.

IPR2022-01289 Page 00129

6.2 Recognizing vowels

Ironically, speech recognition research began with vowel and phoneme detection. Vowels are typically long in duration compared to consonants and are spectrally well-defined so the presence of one can be reliably detected [Rabiner, 1993], but identifying the presence of a specific vowel varies from speaker-to-speaker and is more difficult as was shown in Table 9. In 1959, researchers at MIT Lincoln Laboratories [Forge, et al.] built a detector capable of recognizing a few vowels contained within a /b/-vowel-/t/ pattern. The system used a filter band analyzer and an estimate of time duration. In 1961, researchers in Japan [Suzuki, et al.] built a hardware vowel recognizer for some Japanese vowels which used an elaborate filter bank spectrum analyzer along with logic that connected the weighted outputs of each channel to a vowel-decision circuit where a majority-decision logic scheme was used to identify the spoken vowel. In 1962 another group of researchers in Japan built a hardware phoneme recognizer [Sakai, et al.] which provided gross categorization using a segmenter along with zero-crossing analysis. In 1966, Reddy pioneered research to dynamically track phonemes for continuous speech recognition.

Today's speech recognition systems still rely heavily on vowel recognition to achieve high performance, but the shift towards statistical processing spawned by Harpy in the 1970s pushed researchers to incorporate higher-level knowledge and accept gross uncertainty at the acoustic-phonetic level or ignore it altogether. In concluding, we return our attention to the BeBe System in light of these works.

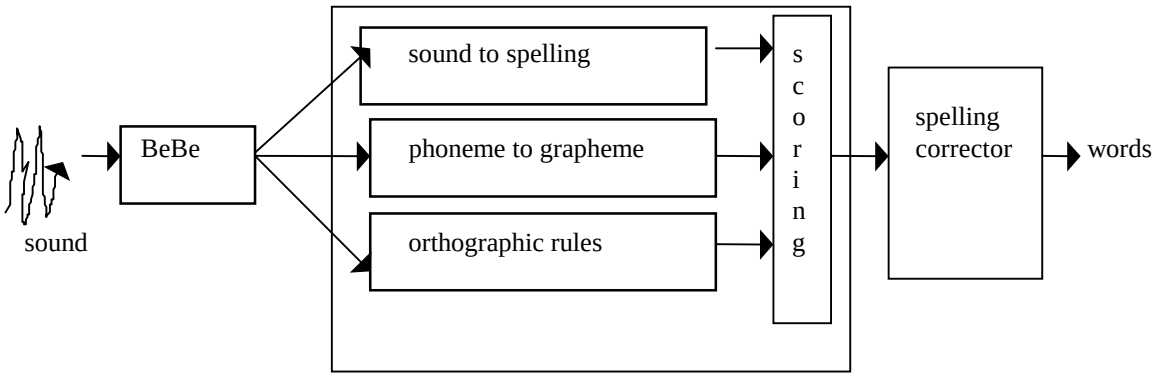


Diagram 3. Block diagram for a speech recognizer that uses the BeBe System as the first stage. The middle stage has a design similar to that of the BeBe System, with three parallel processes, where each process computes spellings from the phoneme-like sequence that's outputted from BeBe. Associated with each of these spellings is a likelihood measure as to how certain the process is that the phoneme sequence represents that spelling. The top few most likely spellings are then sent to a spelling corrector where the most likely words are determined and outputted.

IPR2022-01289 Page 00130

6.3 Assessment and future work

Consider the phoneme detectors for /@/, /R/, /I/, and /U/ discussed earlier in the Implementation and Results sections. $BeBe_0$ used FFT spectrum analysis and weights, time analysis, and a voting scheme to identify phonemes. This is similar to the technology employed by researchers in the 1950s and 1960s for building recognizers, though admittedly $BeBe_0$'s implementation is more powerful and takes advantage of improved computational prowess. Some of its power comes from its simplicity.

Having each detector report a certainty factor eliminates complicated and conflicting hierarchical rules which would otherwise be inevitable as the system would expand to include more detectors. Unlike previous recognizers and current classifiers, where the same procedures are applied to the recognition of all phonemes, detectors in the BeBe architecture are independent and can exploit other methodologies. For example, pitch, spectral tilt, relative formant amplitude and wavelet analysis were not necessary for detecting phonemes in $BeBe_0$ but may prove quite useful in the identification of other phonemes. The BeBe architecture also supports the use of special detectors to identify ambient noises, and perhaps eventually, to isolate different speakers. Assigning one or more detectors to each phoneme allows the recognition algorithms to remain simple and the overall result produces well-modulated abstractions. Complexity in $BeBe_0$ was governed by the required number of multiplications for the FFT which was $O(n \lg n)$, where $n = 2^k$ is the window size. This is constant during operation and spatial complexity is likewise constant.

The phoneme detectors in $BeBe_0$ identified vowels but many of the same spectral tools will be useful in identifying consonants as well [Weinstein, et al., 1975]. Further, the results in Table 9 suggest that machines can detect consonants more accurately than vowels.

Unfortunately, each of the detectors in $BeBe_0$ were hand-coded and this conflicts with the adaptive nature of human auditory perception. We believe however that coding detectors by hand will provide a set of tools from which a representation will emerge that can be used to computationally "evolve" detectors. In such a case, the resulting BeBe System would hear sounds and then build its own detectors, thereby completely emulating the auditory learning behavior described by the Korean language students.

IPR2022-01289 Page 00131

References

[Allen, 1994] Allen, J. How do humans process and recognize speech? *IEEE Transactions on Speech and Audio Processing*: 2(4).

[Baker, 1975] Baker, J. Stochastic modeling for automatic speech understanding. *Speech Recognition*, edited by Raj Reddy. London: Academic Press: 521-541.

[Bourlard, et al., 1996] Bourlard, H., Hermansky, H., and Morgan, N. Towards increasing speech recognition error rates. *Speech Communication*: 18(3).

[Carlson, et al., 1992] Carlson, R. and Glass, J. Vowel classification based on analysis-by-synthesis. *Proceedings of ICSLP*, 575-578.

[Chigier, et al., 1992] Chigier, B. and Leung, H. The effects of signal representations, phonetic classification techniques. *Proceedings of ICSLP*, 97-100.

[Cole, et al., 1980] Cole, R.A., Rudnicky, A.I., Zue, V.W. and Reddy, D.R. Speech as patterns on paper. *Perception and Production of Fluent Speech*, edited by R.A. Cole. Hillsdale: Lawrence Erlbaum Associates: 3-50.

[Cole, et al., 1992] Cole, R. and Muthusamy, Y. Perceptual studies on vowels excised from continuous speech. *Proceedings of ICSLP*, 1091-1094.

[Digalakis, et al., 1993] Digalakis, V., Rohlicek, J., and Ostendorf, M. ML estimation of a stochastic linear system with the EM algorithm and its application to speech recognition. *IEEE Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 431-442.

[Erman, et al., 1980] Erman, L., Hayes-Roth, F., Lesser, V. and Reddy, D.R. The Hearsay-II speech-understanding system: integrating knowledge to resolve uncertainty. *Computing Surveys*, 12(2), 213-251.

[Forgie, et al., 1959] Forgie, J. And Forgie, C. Results obtained from a vowel recognition computer program. *Journal of Acoustic Society of America*, 31(11):1480-1489.

[Hall, 1961] Hall, R.A. *Sound and spelling in English*. Philadelphia.

[House, 1962] House, A.S. On the learning of speechlike vocabularies. *Journal of Verbal and Learning Verbal Behavior*, 1, 133-143.

[Huang, 1991] Huang, C.B. *An acoustic and perceptual study of vowel formant trajectories in American English.* Massachusetts Institute of Technology, Research Laboratory of Electronics: Technical Report 563.

[Jones, 1993] Jones, E.R. *Contemporary College Physics*, 2nd ed., Reading: Addison-Wesley: 426.

[Klatt, 1977] Klatt, D. Review of the ARPA Speech Understanding Project. *Journal of the Acoustic Society of America*, 62, 1345-1366.

[Kornfeld, 1979] Kornfeld, W.A. *Using parallel processing for problem solving.* Massachusetts Institute of Technology, Artificial Intelligence Laboratory: Memo 561.

[Liberman, 1957] Liberman, A.M. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358-368.

[Lippmann, 1996] Lippmann, R. Recognition by humans and machines: miles to go before we sleep. *Speech Communication*: 18(3), 247-248.

[Markowitz, 1996] Markowitz, J. *Using speech recognition.* Upper Saddle River: Prentice-Hall: 129-135

[Meng, et al., 1991] Meng, H. and Zue, V. Signal representation comparison for phonetic classification. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 285-288.

[Peterson and Barney, 1952] Peterson, G.E. and Barney, H.L. Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 24(2), 175-184.

[Pieraccini, et al., 1991] Pieraccini, R., Lee, C., Giachin, E. and Rabiner, L. Complexity reduction in a large vocabulary speech recognizer. *IEEE Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 729-732.

[Rabiner, et al., 1993] Rabiner, L. and Juang, B. *Fundamentals of speech recognition.* Englewood Cliffs: Prentice-Hall: 392.

[Reddy, 1966] Reddy, D. *An approach to computer speech recognition by direct analysis of the speech wave*. Stanford University, Computer Science Department, Technical Report C549.

[Rimmer, 1995] Rimmer, S. *Advanced multimedia programming.* New York: Windcrest/McGraw-Hill.

[Sakai, et al., 1962] Sakai, T. and Doshita, S. The phonetic typewriter, information processing. *Proceedings IFIP Congress.* Munich. Also discussed in [Rabiner, et al., 1993].

[Schmid, 1996] Schmid, P. *Explicit, N-best formant features for segment-based speech recognition*. Oregon Graduate Institute, Department. of Computer Science, Technical Report CSE-96-TH-003.

[Schroeder, et al., 1979] Schroeder, M.R., Atal, B.S., and Hall, J.L. Objective measure of certain speech signal degradations based on masking properties of human auditory perception.

IPR2022-01289 Page 00133

Frontiers of Speech Communication Research, edited by B. Lindblom and S. Ohman. London: Academic Press: 217-229.

[Suzuki, et al., 1961] Suzuki, J. and Nakata, K. Recognition of Japanese vowels – preliminary to the recognition of speech. *Journal Radio Research Laboratory* : 37(8):193-212.

[Sweeney, 1996a] Sweeney, L. Replacing Personally-Identifying Information in Medical Records, the Scrub System. In: Cimino, JJ, ed. Proceedings, *Journal of the American Medical Informatics Association.* Washington, DC: Hanley & Belfus, Inc, 1996:333-337.

[Sweeney, 1996b] Sweeney, L. Automatic acquisition of orthographic rules for recognizing and generating spellings. MIT. AI Working Paper.

[Weinstein, et al. 1975] Weinstein, C., McCandless, S., Modshein, L., and Zue, V. A system for acoustic-phonetic analysis of continuous speech. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23(1), 314-327.

[Zue, 1985] Zue, V. W. The use of speech knowledge in automatic speech recognition. *Proceedings of the IEEE*, 73(11), 1602-1615.

ATTACHMENT

Q

Wayback Machine Search for Sweeney and Thompson report at:

http://publications.csail.mit.edu/lcs/pubs/pdf/MIT-LCS-TR-736.pdf

There are three captures, the earliest of which is July 12, 2010.

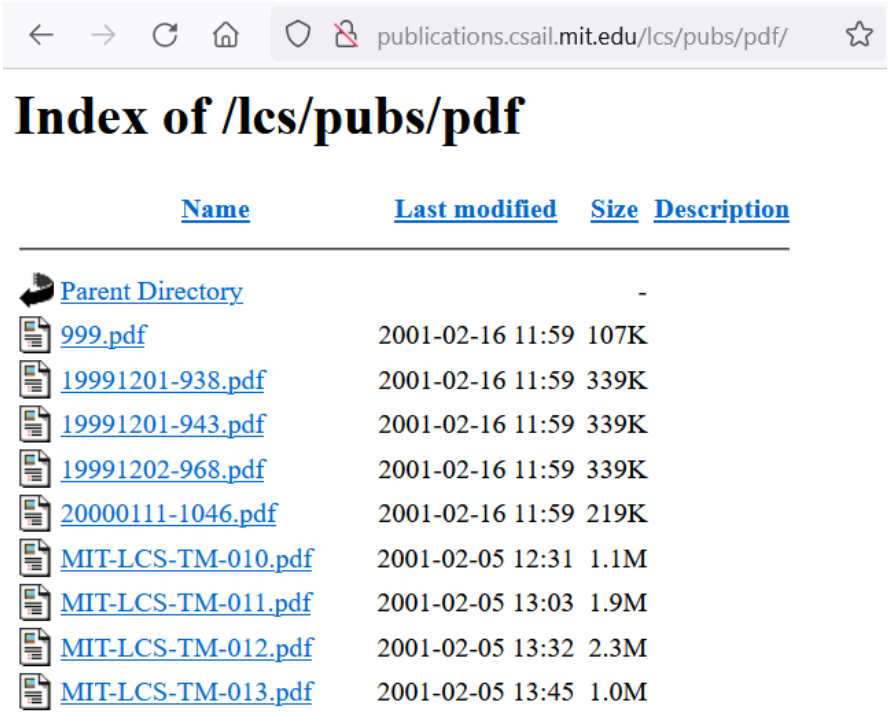


ATTACHMENT

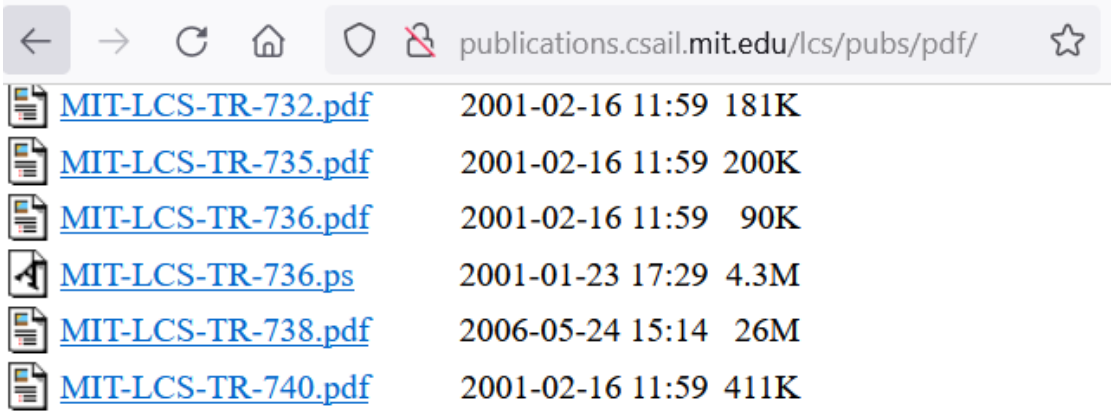
R

Results of search for: http://publications.csail.mit.edu/lcs/pubs/pdf/

1. Index of all files at lcs/pubs/pdf and the date each file was last modified.



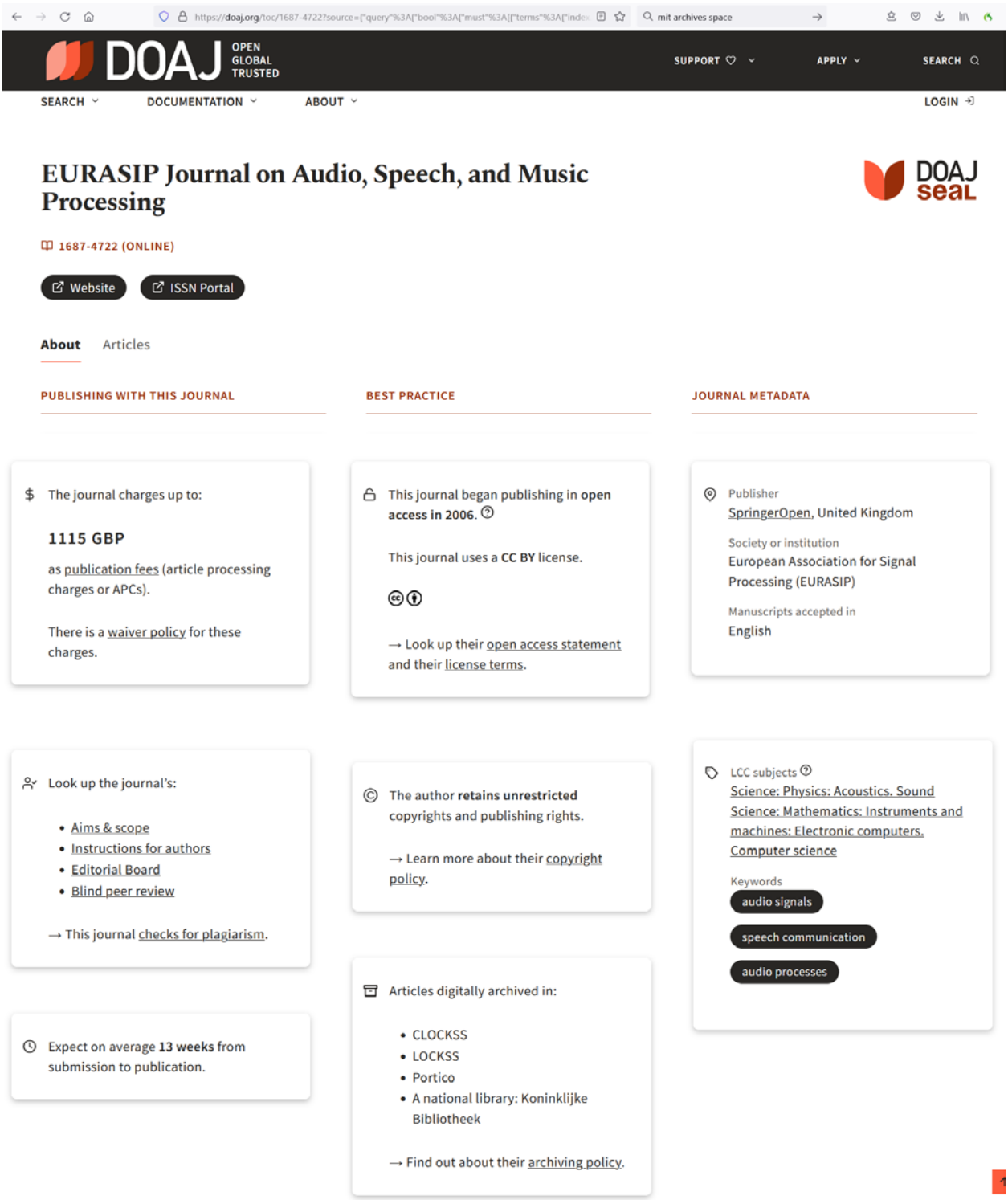
2. “MIT-LCS-TR-736.pdf”, the Sweeney and Thompson report file, was last modified on Feb. 16, 2001.



ATTACHMENT

S

Information about *EURASIP Journal on Audio, Speech, and Music Processing* from the Directory of Open Access Journals (DOAJ). Available: https://doaj.org/toc/1687-4722



Deposit policy with:

- Sherpa/Romeo

Permanent article identifier:

- DOI

ATTACHMENT

T

Welcome, Guest. Sign in to your institution for full access to your library's resources.

EURASIP journal on audio, speech, and music processing [electronic resource].

Language:	English
Publication:	Springer Nature Complete **Journals**; DOAJ: Directory of Open Access **Journals**; EBSCO Open Access **Journals**
Publication Information:	New York, NY : Hindawi Pub. Corp., 2007- <2011> : Heidelberg : Sp
Publication Date:	2007
Publication Type:	Periodical
Document Type:	Periodical
Subject Terms:	Sound -- Recording **and** reproducing -- Digital techniques -- Periodicals Computer sound **processing** -- Periodicals Computer sound **processing** Sound -- Recording **and** reproducing -- Digital techniques Periodicals Electronic **journal** Electronic **journals** Electrical Engineering
Notes:	Title from title screen (Hindawi, viewed Dec. 5, 2006). Articles continually posted as accepted in "Articles in press" section; articles also browseable by quarter of acceptance. Text (electronic **journal**). Special issues accompany v. Mode of access: World Wide Web. Publication of the European Association for **Speech**, Signal, **and** Image **Processing**.
Other Authors:	European Association for **Speech**, Signal **and** Image **Processing**. Hindawi Publishing Corporation.
Other Titles:	European Association for **Speech**, Signal, **and** Image **Processing journal on audio, speech, and music processing**. **EURASIP** JASMP. **Audio**, **speech**, **and music**

processing.
EURASIP Journal on Audio, **Speech**, **and Music Processing**
EURASIP journal on audio, **speech**, **and music processing** (Online)

ISSN: 1687-4722

Publication Frequency: Annual

Online Access: This title is available in full-text from 10/2006 to present in Springer Nature Complete Journals. On Campus or Off Campus Access, CLICK HERE.
This title is available in full-text from 2006 to present in DOAJ: Directory of Open Access Journals. On Campus or Off Campus Access, CLICK HERE.
This title is available in full-text from 2006 to present in EBSCO Open Access Journals. On Campus or Off Campus Access, CLICK HERE.

Accession Number: ua.7643151

Database: University of Alabama Libraries' Classic Catalog

Request This

EBSCO Connect | Privacy Policy | A/B Testing | Terms of Use | Copyright | Cookie Policy | Contact Us

powered by EBSCO*host*

© 2022 EBSCO Industries, Inc. All rights reserved.

ATTACHMENT

U

Springer Open

Search Get published Explore Journals Books About Login

Menu

EURASIP Journal on Audio, Speech, and Music Processing

About Articles Submission Guidelines

Download PDF

Research Article | Open Access |
Published: 04 October 2007

Voice-to-Phoneme Conversion Algorithms for Voice-Tag Applications in Embedded Platforms

YanMing Cheng, Changxue Ma & Lynette Melnar

EURASIP Journal on Audio, Speech, and Music Processing
2008, Article number: 568737 (2007)

3598 Accesses | Metrics

Abstract

We describe two voice-to-phoneme conversion algorithms for speaker-independent voice-tag creation specifically targeted at applications on embedded platforms. These algorithms (*batch mode* and *sequential*) are compared in speech recognition experiments where they are first applied in a same-language context in which both acoustic model

IPR2022-01289 Page 00146

training and voice-tag creation and application are performed on the same language. Then, their performance is tested in a cross-language setting where the acoustic models are trained on a particular source language while the voice-tags are created and applied on a different target language. In the same-language environment, both algorithms either perform comparably to or significantly better than the baseline where utterances are manually transcribed by a phonetician. In the cross-language context, the voice-tag performances vary depending on the source-target language pair, with the variation reflecting predicted phonological similarity between the source and target languages. Among the most similar languages, performance nears that of the native-trained models and surpasses the native reference baseline.

1. Introduction

A voice-tag (or name-tag) application converts human speech utterances into an abstract representation which is then utilized to recognize (or classify) speech in subsequent uses. The voice-tag application is the first widely deployed speech recognition application that uses technologies like Dynamic Time Warping (DTW) or HMMs in embedded platforms such as in mobile devices.

Traditionally, HMMs are directly used as abstract speech representations in voice-tag applications. This approach has enjoyed considerable success for several reasons. First, the approach is language-

IPR2022-01289 Page 00147

independent so it is not restricted to any particular language. With the globalization of mobile devices, this feature is imperative as it allows for speaker-dependent speech recognition for potentially any language or dialect. Second, the HMM-based voice-tag technology achieves high speech recognition accuracy while maintaining a low CPU requirement. The storage of one HMM per voice-tag, however, is rather significant for many embedded systems, especially for low-tier ones. Only as long as storage is kept under the memory budget of an embedded system by limiting the number of voice-tags, is the HMM-based voice-tag strategy acceptable. Usually, two or three dozen voice-tags are recommended for low-tier embedded systems, while high-tier embedded systems can support a greater number. Nevertheless, interest in constraining the overall cost of embedded platforms limits the number of voice-tags in practice. Finally, the HMM-based voice-tag has been successful because it is speaker-dependent and convenient for the user to create during the enrollment phase. A typical enrollment session in a speaker-dependent context requires only a few sample utterances to train a voice-tag HMM that captures both the speech abstraction and the speaker characteristics.

Today, speaker-independent and phoneme HMM-based speech recognizers are being included in mobile devices, and voice-tag technologies are mature enough to leverage the existing computational resources and algorithms from the speaker-independent speech recognizer for further

IPR2022-01289 Page 00148

efficiency. A name dialing application can recognize thousands of names downloaded from a phonebook via grapheme-to-phoneme conversion, and voice-tag technology is a convenient way of dynamically extending the voice-enabled phonebook. In this type of application, voice-tag entries and phonetically transcribed name entries are jointly used in a speaker-independent context. Limiting voice-tags to two or three dozen in this scenario may no longer be practical, though extending the number significantly in a traditional HMM-based voice-tag application could easily surpass the maximum memory consumption threshold for low-tier embedded platforms. Given this, the speaker-dependency of the traditional approach may actually prevent the combined use of the voice-tag HMM technology and phonetically transcribed name entries.

Assuming that a set of speaker-independent HMMs already resides in an embedded platform, it is natural to think of utilizing a phonetic representation (phoneme strings or lattices) created from sample utterances as the abstract representation of a voice-tag, that is, voice-to-phoneme conversion. The phonetic representation of a voice-tag can be stored as cheaply as storing a name entry with a phonetic transcription, and it can be readily used in conjunction with the name entries obtained via grapheme-to-phoneme conversion, as long as such a voice-tag is speaker-independent. Voice-to-phoneme conversion, then, enhances speech recognition capability in embedded platforms by greatly extending recognition coverage.

IPR2022-01289 Page 00149

As mentioned, a practical concern of voice-tag technology is user acceptability. Extending recognition coverage from dozens to hundreds of entries without maintaining or improving recognition and user convenience is not a viable approach. User acceptability mandates that the number of sample utterances required to create a voice-tag during enrollment be minimal, with one sample utterance being most favorable. However, in a speech recognition application with a large number of voice-tags, the recognition accuracy of each voice-tag tends to increase as its number of associated sample utterances increases. So, in order to achieve acceptable performance, more than one sample utterance is typically required during enrollment. Generally, a compromise of two to three sample utterances is usually considered acceptable.

Voice-to-phoneme conversion has been investigated in modeling pronunciation variations for speech recognition ([1, 2]), spoken document retrieval ([3, 4]) and word spotting ([5]) with noteworthy success. However, optimal conversion in the sense of maximum likelihood presented in these prior works requires prohibitively high computation, which prevents their direct deployment to an embedded platform. This crucial problem was resolved in [6], where we introduced our batch mode and sequential voice-to-phoneme conversion algorithms for speaker-independent voice-tag creation. In Section 2 we describe the batch mode voice-to-phoneme conversion algorithm in particular and show how it meets the criteria of low computational complexity

IPR2022-01289 Page 00150

and memory consumption for a voice-tag application in embedded platforms. In Section 3 we review the sequential voice-to-phoneme algorithm which both addresses user convenience during voice-tag enrollment and improves recognition accuracy. In Section 4, we discuss the experiment conditions and results for both the batch mode and sequential algorithms in a same-language context.

Finally, a legitimate concern confronting any voice-tag approach is its language extensibility. Increasingly, the task of extending a technology to a new language must consider the potential lack of sufficient target-language resources on which to train the acoustic models. An obvious strategy is to use language resources from a resource-sufficient source language to recognize a target language for which little or no speech data is assumed. Several studies have in fact explored the effectiveness of the cross-language application of phoneme acoustic models in speaker-independent speech recognition (see [7–10]). In [11], we demonstrated the cross-language effectiveness of the batch mode and sequential voice-to-phoneme conversion algorithms. Section 5 documents the cross-language voice-tag experiments and provides the results in comparison with that of those achieved in the same-language context. Here, we analyze the cross-language results in terms of predicted global phonological distance between the source and target languages. Finally, we share some concluding remarks in Section 6.

2. Batch-Mode Voice-to-Phoneme

IPR2022-01289 Page 00151

Conversion

The principle idea of batch mode creation is to use a feature-based combination (here, DTW) collapsing M sample utterances (hereinafter *samples*) into a single "average" utterance. The expectation is that this "average" utterance will preserve what is common in all of the constituent samples while neutralizing their peculiarities. As mentioned in the previous section, the number of enrollment samples directly affects voice-tag accuracy performance in speech recognition. The greater the number of samples during the enrollment phase, the better the performance is expected to be.

Let us consider that there are M samples, $\mathbf{X}_m \ (m \in [1, M])$, available to a voice-tag in batch mode. $\mathbf{X}_m$ is a sequence of feature vectors corresponding to a single sample. For the purpose of this discussion, we do not distinguish between a sequence of feature vectors and a sample utterance in the remaining part of this paper. The objective here is to find the N-best phonetic strings, $\mathbf{P}_n \ (n \in [1, N])$, following an optimization criterion.

In prior works describing a batch mode voice-to-phoneme conversion method, the tree-trellis N-best search algorithm [12] is applied to find the optimal phonetic strings in the maximum likelihood sense [1, 2]. In [1, 2], the tree-trellis algorithm is modified to include a backward, time asynchronous, tree search. First, this modified tree-trellis search algorithm produces a tree of partial phonetic hypotheses for each sample using conventional time-synchronized

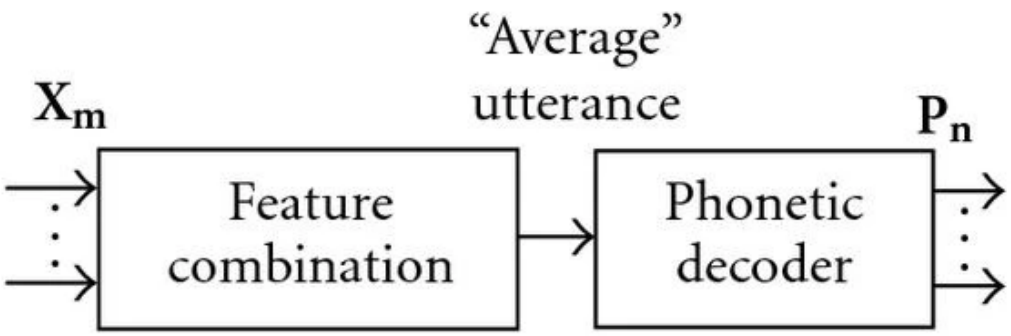
IPR2022-01289 Page 00152

Viterbi decoding and a phoneme-loop grammar in the forward direction. The M trees of phonetic hypotheses of the samples are used jointly to estimate the admissible partial likelihoods from each node in the grammar to the start node of the grammar. Then, utilizing the admissible likelihood of partial hypotheses, an A^*-search in the backward direction is used to retrieve the N best phonetic hypotheses, which maximize the likelihood. The modified tree-trellis algorithm generally falls into the probability combination algorithm category. Because this algorithm requires storing M trees of partial hypotheses simultaneously, with each tree being rather large in storage, it is expensive in terms of memory consumption. Furthermore, the complexity of the A^* search increases significantly as the number of samples increases. Therefore, this probability combination algorithm is not very attractive for deployment in mobile devices.

To meet embedded platform requirements, we use a feature-based combination algorithm to perform voice-to-phoneme conversion to create a voice-tag. By combining the M sample utterances of different lengths into a single "average" utterance, a simple phonetic decoder, e.g., the original form of the tree-trellis search algorithm with a looped phoneme grammar, can be used to obtain N best phonetic strings per voice-tag with minimum memory and computational consumption. Figure 1 depicts the system.

Figure 1

IPR2022-01289 Page 00153



Voice-to-phoneme conversion in batch mode.

DTW is of particular interest to us because of the memory and computation efficiency of its implementation in an embedded platform. Many embedded platforms have a DTW library specially tuned to the target hardware. Given two utterances, $\mathbf{X}_i$ and $\mathbf{X}_j$ $(i \neq j$ and $i,j \in [1,M])$, a trellis can be formed with $\mathbf{X}_i$ and $\mathbf{X}_j$ being horizontal and vertical axes, respectively. Using a Euclidean distance and DTW algorithm, the best path can be derived, where "best path" is defined as the lowest accumulative distance from the lower-left corner to the upper-right corner of the trellis. A new utterance $\mathbf{X}_{i,j}$ can be formed along the best path of the trellis, $\mathbf{X}_{i,j} = \mathbf{X}_i \oplus \mathbf{X}_j$, where $\oplus$ is denoted as the DTW operator. The length L of the new utterance is the length of the best path.

Let:

$$\mathbf{X}_{i,j} = \{\mathbf{x}_{i,j}(0), \ldots, \mathbf{x}_{i,j}(t), \ldots, \mathbf{x}_{i,j}(L_{i,j} - 1)\},$$

$$\mathbf{X}_i = \{\mathbf{x}_i \mathbf{X}_i = \{\mathbf{x}_i(0), \ldots, \mathbf{x}_i(c), \ldots, \mathbf{x}_i(L_i - 1)\}$$ and

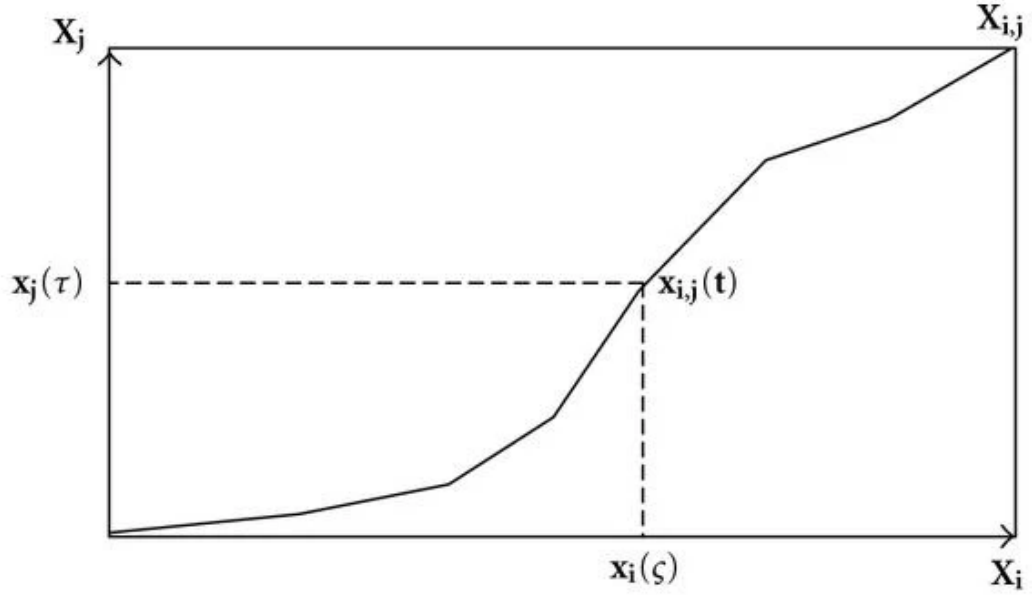
IPR2022-01289 Page 00154

$$\mathbf{X}_j = \{\mathbf{x}_j(0), \ldots, \mathbf{x}_j(\tau), \ldots, \mathbf{x}_j(L_j - 1)\},$$

where t, ς, τ are frame indices.

We define $\mathbf{x}_{i,j}(t) = (\mathbf{x}_i(\varsigma) + \mathbf{x}_j(\tau))/2$, where t is the position on the best path aligned to the ςth frame of $\mathbf{X}_i$ and the τth frame of $\mathbf{X}_j$ according to the DTW algorithm. Figure 2 sketches the feature combination of two samples.

Figure 2



DTW-based feature combination of two sample utterances.

Given M samples $\mathbf{X}_1 \ldots \mathbf{X}_M$ and the feature combination algorithm of two utterances, there are many possible ways of producing the final "average" utterance. Through experimentation we have found that they all

IPR2022-01289 Page 00155

achieve statistically similar speech recognition performances. Therefore, we define the "average" utterance as the cumulative operation of the DTW-based feature combination:

$$\mathbf{X}_{1,2,3,\ldots,M} = (\cdots((\mathbf{X}_1 \oplus \mathbf{X}_2) \oplus \mathbf{X}_3) \cdots \oplus \mathbf{X}_M). \quad (1)$$

The cumulative operation provides a storage advantage for the embedded system.Independent of M, only two utterances need to be stored at any instance: the intermediate "average" utterance and the next new sample utterance. The computation complexity of the batch-mode voice-to-phoneme conversion is the M-1 DTW operation, one tree decoding with a phoneme-loop grammar and a trellis decoding for N-best phonemic strings. As a comparison, the approach in [1, 2] needs to store at least one sample utterance and M rather large phonemic trees; furthermore, it requires the computation of M tree decoding with the phoneme loop grammar and a trellis decoding. Thus, the proposed batch mode algorithm is more suited for an embedded system.

3. Sequential Voice-Tag Creation

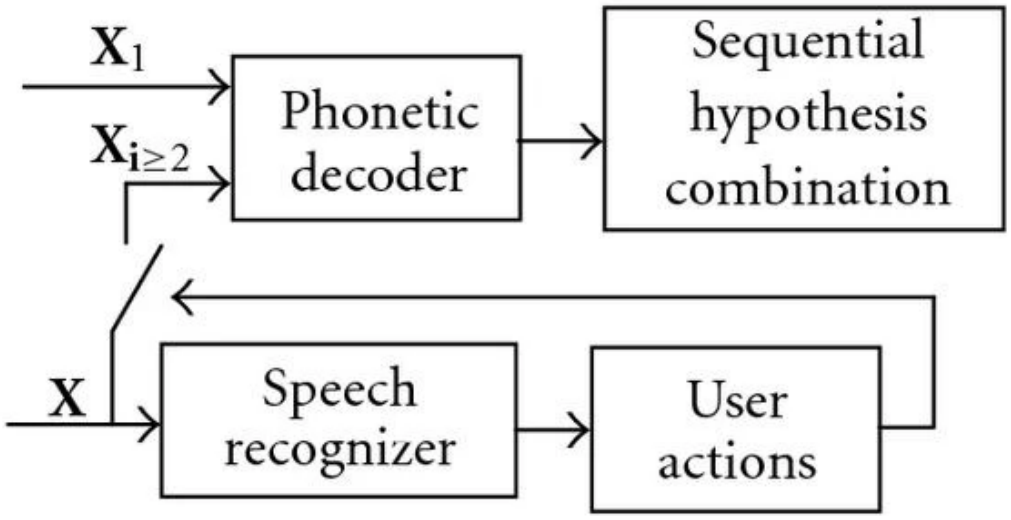
Sequential voice-tag creation is based on the hypothesis combination of the outputs of a phonetic decoder of M samples. In this approach, only one sample per voice-tag is required to create N initial seed phonetic strings, $\mathbf{P}_n$, using a phonetic decoder. The phonetic decoder used here is the same as

IPR2022-01289 Page 00156

described in the previous section. If good phonetic coverage is exhibited by the phonetic decoder (i.e., good phonetic robustness of the trained HMMs), with initial seed phonetic strings the recognition performance of voice-tags is usually acceptable, though not maximized. Each time a voice-tag is successfully utilized (a positive confirmation of the speech recognition result is detected and the corresponding action is implemented—e.g., the call is made), the utterance is reused as another sample to produce additional N phonetic strings to update the seed phonetic strings of the voice-tag through performing hypothesis combination. This update can be performed repeatedly until a maximum performance is reached. Figure 3 sketches this system.

IPR2022-01289 Page 00157

Figure 3



Sequential combination of hypothetical results of a phonetic decoder.

The objective of this method is to discover a sequential hypothesis combination algorithm that leads to maximum performance. We use a hypothesis combination based on a consensus hierarchy displayed in the best phonetic strings of samples. The consensus hierarchy is expressed numerically in a phoneme n-gram histogram (typically a monogram or bigram is used).

3.1. Hierarchy of Phonetic Hypotheses

To introduce the notion of "hierarchy of phonetic hypotheses," let us begin by showing a few examples. Suppose we have three samples of the name "Austin." The manual phonetic transcription of this name is /= s t I n/. The following list shows the best string

IPR2022-01289 Page 00158

obtained by the phonetic decoder for each sample:

$P_1(X_1)$: f pau ɔ s I n d,
$P_1(X_2)$: k pau ɔ f I n d,
$P_1(X_3)$: pau ɔ f I n.

(2)

The next list shows the three best phonetic strings of the first sample obtained by the phonetic decoder:

$P_1(X_1)$: f pau ɔ s I n d,
$P_2(X_1)$: t pau ɔ s I n d,
$P_3(X_1)$: f pau ɔ s I n z.

(3)

Examining the results of the phonetic decoder, we observe that there are some phonemes that are very stable across the samples and the hypotheses; further, these phonemes tend to correspond to identical phonemes in the manual transcription. It is also observed that other phonemes are quite unstable across the samples. These derive from the peculiarities of each sample and the weak constraint of the phoneme-loop grammar. Similar observations are also made in [13], where stable phonemes in particular are termed the "consensus" of the phonetic string. Since we are interested in embedded voice-tag speech recognition applications in potentially diverse environments, we investigated phoneme stability in both favorable and unfavorable conditions. Our investigation shows that in a noisy environment some phonemes still remain stable while others become less stable compared to a more quiet environment. In

IPR2022-01289 Page 00159

general however, a hierarchical phonetic structure for each voice-tag can be easily detected, independent of the environment. At the top level of the hierarchy are the most stable phonemes that reflect the consensus of all instances of the voice-tag abstraction. The phonemes at the middle level of the hierarchy are less stable but are observed in the majority of voice-tag instances. The lowest level in the hierarchy includes the random phonemes, which must be either discarded or minimized in importance during voice-to-phoneme conversion.

3.2. Phoneme N-Gram Histogram-Based Sequential Hypothesis Selection

To describe the hierarchy of a voice-tag abstraction, we utilize a phoneme n-gram histogram. The high frequency phoneme n-grams correspond to "consensus" phonemes, the median frequency to "majority" phonemes and the low frequency to "random" phonemes of a voice-tag. In this approach, it is straightforward to estimate sequentially the n-gram histogram via a cumulative operation. e.g., one can use the well-known *relative entropy* measure [14] to compare two histograms. Another favorable attribute of this approach is that the n-gram histogram can be stored efficiently. For instance, given a phonetic string of length L, there are at most L monograms, $L+1$ bigrams or $L+2$ trigrams without counting zero frequency n-grams. In practice, the n-gram histogram of a voice-tag is estimated based only on the best phonetic strings of the previous utterances and the current utterance of the voice-tag. We ignore all but the best phonetic string of any

IPR2022-01289 Page 00160

utterance in the histogram estimation because the best phonetic string differs from the second best or third best by only one phoneme by definition. This "defined" difference may not be helpful in revealing the majority and random phonemes in a statistical manner; it may even skew the estimated histogram.

The Sequential Hypothesis Combination Algorithm Is Provided Below:

Enrollment (or initialization): Use one sample per voice-tag to create N phonetic strings via a phonetic decoder as the current voice-tag; use the best phonetic string to create the phoneme n-gram histogram for the voice-tag.

Step 1. Given a new sample of a voice-tag, create N new phonetic strings (via the phonetic decoder); update the phoneme n-gram histogram of the voice-tag with the best phonetic string of the new sample.

Step 2. Estimate a phoneme n-gram histogram for each phonetic string for N current and N new phonetic strings of the voice-tag.

Step 3. Compare the phoneme n-gram histogram of the voice-tag with that of each phonetic string using a distance metric, such as *relative entropy* measure; select N phonetic strings, the histograms of which are closest to the histogram of the voice-tag histogram, as the updated voice-tag representation.

Step 4. Repeat steps 1–3 if a new sample is available.

IPR2022-01289 Page 00161

4. Same-Language Experiments

The database selected for this evaluation is a Motorola-internal American English name database which contains a mixture of both landline and wireless calls. The database consists of spoken proper names of variable length. These names are representational of a cross-section of the United States and predominantly include real-use names of European, South Asian, and East Asian origin. No effort was made to control the length of the average name utterance, and no bias was provided toward names with greater length. Most callers speak either Standard American English or Northern Inland English, though there are a number of other English speech varieties represented as well, including foreign-accented English (especially Chinese and Indian language accents). The database is divided into voice-tag creation and evaluation sets where the creation set has 85 name entries corresponding to 85 voice-tags, and each name entry comprises three samples spoken by a single speaker in different sessions. Thus the creation set is speaker-dependent. The purpose of designing a speaker-dependent creation set is that we expect that any given voice-tag will be created by a single user in real applications and not by multiple users. The evaluation set contains 684 utterances of the 85 name entries. Most speakers of a name entry in the evaluation set are different from the speaker of the same name entry in the creation set, though, due to name data limitations, a few speakers are the same for both sets. In general, however, our evaluation set is speaker-independent.

IPR2022-01289 Page 00162

We use the ETSI advanced front-end standard for distributed speech recognition [15]. This front end generates a feature vector of 39 dimensions per frame and the feature vector contains 12 MFCC plus energy and their delta and acceleration coefficients. The phonetic decoder is the MLite++ ASR search engine, a Motorola proprietary HMM-based search engine for embedded platforms, with a phoneme loop grammar. The search engine uses both context-independent (CI) and context-dependent (CD) sub word and speaker-independent HMMs, which were trained on a much larger speaker-independent American English database than the above spoken name database.

For comparative purposes, the 85 name entries were carefully transcribed by a phonetician. The number of transcriptions per name entry is varied from 1 to many (due to pronunciation differences associated with the distinct speech varieties), with an average of 3.89 per entry. Using these reference transcriptions and a word-bank grammar (i.e., a list of words with equal probability) of the 85 name entries, the baseline word accuracies of 91.67% and 92.69% are obtained on the speaker-independent test set of the spoken name database with CI and CD HMMs, respectively.

4.1. Same-Language Experiments Using Batch Mode Voice-to-Phoneme Conversion

To illustrate the effectiveness of the DTW-based feature combination algorithm, consider the following phonetic strings generated from (i) a single

IPR2022-01289 Page 00163

sample, (ii) an "average" of two samples and (iii) an "average" of three samples of the name "Larry Votta" (where "mn" signifies mouth noise). The manual reference transcription of this name is /l ɛ r iʲ v ɔ t *a*/.

In general, those phonetic strings generated from more samples have more agreement with the manual transcription than those generated with fewer samples. In particular, the averaged strings tend to eliminate the "random" phonemes (like /ŋ/, /ɐ/, /j/, /z/, and /n/,) and preserve the "consensus" and "majority" phonemes (like /l/, /eʲ/, /r/, /iʲ/, and /ɔ/)—which tend to be associated with the manual transcription. Therefore, the DTW-based feature combination does preserve the commonality of the samples, which is expected to be the abstraction of a voice-tag. It is worthwhile to note that the "new" phoneme, which may not necessarily be in original sample utterance, can be generated by the phonetic decoder from the "average" utterance. For instance, /v/ in $P_1(X_{1,2,3})$ is not part of $P_1(X_1)$, $P_1(X_2)$ or $P_1(X_3)$.

Table 2 shows the speech recognition performances of the batch mode voice-to-phoneme conversion. Word accuracy is derived from the test set of the name database where a voice-tag is created with three phonetic strings from a single sample, an "average" of two samples, and an "average" of all three samples of the voice-tag in the speaker-dependent training set.

IPR2022-01289 Page 00164

Table 1

Table 2 Voice-tag word accuracy obtained by batch mode voice-to-phoneme conversion.

As expected, the performance increases when the number of averaged samples per voice-tag increases. When three samples are used the performance is very close to the baseline performance. It is interesting to note that the CI HMMs yield a better performance than the CD HMMs. Further investigation reveals that when varying the ASR search engine configuration, such as the penalty at phoneme boundaries, the performance of the CI HMMs degrades drastically while that of the CD HMMs remains consistent.

4.2. Same-Language Experiments Using Sequential Voice-to-Phoneme Conversion

In this section we only investigate the phoneme monogram and bigram voice-tag histograms for the sequential voice-to-phoneme conversion. Tables 3 and 4 show the speech recognition performance on the test set of the spoken name database with up to three hypothetical phonetic strings generated per each sample via the phonetic decoder. In these results the voice-tags are created sequentially from the speaker-dependent training set of the name database with both monogram and bigram phoneme

IPR2022-01289 Page 00165

histograms.

Table 3 Voice-tag word accuracy obtained by sequential voice-to-phoneme conversion with bigram phoneme histogram.

Table 4 Voice-tag word accuracy obtained by sequential voice-to-phoneme conversion with monogram/bigram phoneme histogram and CD HMMs.

In these experiments, it is observed that word accuracy generally increases when two or more samples are used: CD HMMs outperform both the CI HMMs and the CD HMMS derived from manual transcriptions. It is also noted that both monogram and bigram phoneme histograms yield similar performances depending on the number of samples used.

In order to understand the implication of the number of hypothetical phoneme strings generated per sample, in Table 5 we show the performance results while varying this number:

Table 5 Word accuracy obtained by sequential voice-to-phoneme conversion considering hypothetical phoneme string number.

IPR2022-01289 Page 00166

This table shows the word accuracies obtained with CD HMMs, three samples per voice-tag, a bigram phoneme histogram, which is estimated based on the best hypothetical phoneme string of each sample, and 1–7 hypothetical phoneme strings per sample. The best result is 96.1% word accuracy, which is achieved with as little as four hypothetical phoneme strings per sample. Considering both user convenience and recognition performance, we suggest that 3 or 4 hypothetical phoneme strings per sample might be optimal for sequential voice-to-phoneme conversion for voice-tag applications.

5. Cross-Language Experiments

In practice, evaluating cross-language performance is complex and poses distinct challenges to same-language performance evaluation. In general, cross-language evaluation can be approached by two principle strategies. One strategy creates voice-tags in several target languages by using language resources, such as HMMs and a looped phoneme grammar, from a single source language. The weakness of this strategy is that it is difficult to normalize the linguistic and acoustic differences across the target languages, a necessary step in creating an evaluation database. The other strategy creates voice-tags in a single target language by using language resources from several distinct source languages. The weakness of this strategy is that language resources differ significantly and it cannot be expected that each source language will be trained with the same amount and type of data. Because we can compare

IPR2022-01289 Page 00167

our training data in terms of quantity and type, we opted to pursue the second strategy for the cross-language experiments presented here.

We selected seven languages as source languages: British English (en-GB), German (de-DE), French (fr-FR), Latin American Spanish (es–LatAm), Brazilian Portuguese (pt-BR), Mandarin (zh-CN-Mand) and Japanese (ja-JP). For each of the source languages, we have sufficient data and linguistic coverage to train generic CD HMMs. The phoneme loop grammar of each source language is constructed from the phoneme set of that language.

Since we used American English in the same-language sequential and batch mode voice-to-phoneme conversion experiments above, and thus have these results for comparison, we selected American English as the target language in the following cross-language experiments. For these, we use the same name database, phonetic decoder, and baseline that we used in the same-language experiments.

5.1. Cross-Language Experiments Using Batch Mode and Sequential Voice-to-Phoneme Conversion

In this investigation, the individual cross-language voice-tag recognition performances are compared to both the same-language results and to each other. To do the latter, a phonological similarity study is conducted between the target language (American English) and each of the selected evaluation languages, the prediction being that cross-language performance would correlate to the relative

IPR2022-01289 Page 00168

phonological similarity of the source languages to the target language. We use a pronunciation dictionary as each language's phonological description in order to ensure task independence; because each language's pronunciations are transcribed in a language-independent notation system (similar to the International Phonetic Alphabet), cross-language comparison is possible [16]. Phoneme-bigram (biphoneme) probabilities collected from each dictionary are used as the numeric expression of the phonological characteristics of the corresponding language. The distance between the biphoneme probabilities of each source language and that of the target language is then measured. This metric thus explicitly provides a biphoneme inventory and phonotactic sequence importance. It also implicitly incorporates phoneme inventory and phonological complexity information. Using this method, the distance score is an objective indication of phonological similarity in the source-target language pair, where the smaller the distance value between the languages, the more similar the pair (see [7] for an in-depth discussion of this biphoneme distribution distance).

The languages that we use in these evaluations are from four language groups defined by genetic relation: (i) Germanic: en-US, en-GB, and de-DE; (ii) Romance: fr-FR, pt-BR, es-LatAm; (iii) Sinitic: zh-CN-Mand and (iv) Japonic: ja-JP. In general, it is expected that closely related languages and contact languages (languages spoken by people in close contact with speakers of the target language [17]),

IPR2022-01289 Page 00169

will exhibit greatest phonological similarity. The distance scores relative to American English are provided in the last column of Table 6. Note that the Germanic languages are measured to be the most similar to American English. In particular, the British dialect of English is least distant to American English, and German, the only other Germanic language in the evaluation set, is next. German is followed by French in phonological distance, and French and English are languages with centuries of close contact and linguistic exchange.

Table 6 Word accuracies of voice-tag recognition with batch mode and sequential voice-tag creations in cross-language experiments.

This preliminary study thus both substantiates in a quantitative way linguistic phonological similarity assumptions and provides a reference from which to evaluate our results. Based on this study, it is our expectation that cross-language voice-tag application performance will be degraded relative to the voice-tag application performance in the same-language setting, and that the severity of the degradation will be a function of phonological similarity.

Table 6 also shows the cross-language voice-tag application performances of the sequential and batch mode voice-to-phoneme conversion algorithms, where the acoustic models are trained on the seven evaluation languages while the voice-tags are created

IPR2022-01289 Page 00170

and applied on American English, a distinct target language. For reference, we also include the American English HMM performance as a baseline.

Apart from the exceptional performance of Mandarin using the sequential phoneme conversion algorithm, the performances generally adhere to the target-source language pair similarity scores identified above. Voice-tag recognition with British English-trained HMMs achieve a word accuracy of 91.37% and recognition with German-trained HMMs realize 90.5%. The higher-than-expected performance rate of Mandarin may be due to some correspondences between the American English and Mandarin databases. The American English database consists of a minority of second language speakers of English, especially native Chinese and Indian speakers.

Thus, the utterances used to train the American English models include some Mandarin-language pronunciations of English words. Secondly, the Mandarin models are embedded with a significant amount of English material (English loan words, e.g.), reflecting a modern reality of language use in China.

The cross-language evaluations show significant performance differences between the two voice-creation algorithms across all of the evaluated languages. The differences are in accordance with our observation in the same-language evaluation. Although there are degradations, the performances of sequential voice-tag creation with HMMs trained on

IPR2022-01289 Page 00171

the languages most phonologically similar to American English are very close to the reference performance (92.69%) where the phonetic strings of voice-tags were transcribed manually by an expert.

6. Discussion and Conclusions

We presented two voice-to-phoneme conversion algorithms, each of which utilizes a phonetic decoder and speaker-independent HMMs to create speaker-independent voice-tags. However, these two algorithms employ radically different approaches for sample combination. It is difficult to theoretically compare the algorithms' creation process complexities, though we have observed that the creation processes of both algorithms require similar computational resources (CPU and RAM). The voice-tag created by both algorithms is a set of phonetic strings that require very low storage, making them suitable for embedded platforms. So, for a voice-tag with N phonetic strings of an average length L, a voice-tag requires N times L bytes to store phonetic strings. For continuous improvement of voice-tag representation, the sequential creation algorithm retains a phoneme n-gram histogram per voice-tag, which requires approximately $2L$ bytes.In a typical case where $N=3$ and $L=7$, 21 bytes are needed for each voice-tag created by the batch-mode algorithm, while the sequential creation algorithm requires 35 bytes for each voice-tag. Both algorithms are shown to be effective in voice-tag applications, as they yield speech recognition performances either comparable to or exceeding a manual reference in same-language

IPR2022-01289 Page 00172

experiments.

In the batch mode voice-to-phoneme conversion algorithm, we focused on preserving the input feature vector commonality among multiple samples as a voice-tag abstraction by developing a feature combination strategy. In the sequential voice-to-phoneme conversion approach, we investigated the hierarchy of phonetic consensus buried in the hypothetical phonetic strings of multiple example utterances. We used an n-gram phonetic histogram accumulated sequentially to describe the hierarchy and to select the most relevant hypothetical phoneme strings to represent a voice-tag.

We demonstrated that the voice-to-phoneme conversion algorithms are not only applicable in a same-language environment, but may also be used in a cross-language setting without significant degradation. For the cross-language experiments, we used a distance metric to show that performance results associated with HMMs trained on languages phonologically similar to the target language tend to be better than results achieved with less similar languages, such that performance degradation is a function of source-target language similarity, providing database characteristics, such as intralingual speech varieties and borrowings, are considered. Our experiments suggest that a cross-language application of a voice-to-phoneme conversion algorithm is a viable solution to voice-tag recognition for resource-poor languages and dialects. We believe this has important consequences given

IPR2022-01289 Page 00173

the globalization of mobile devices and the subsequent demand to provide voice technology in new markets.

References

1. Holter T, Svendsen T: **Maximum likelihood modeling of pronunciation variation.** *Speech Communication* 1999,**29**(2–4):177-191.

2. Svendsen T, Soong FK, Purnhagen H: **Optimizing baseforms for HMM-based speech recognition.** *Proceedings of the 4th European Conference on Speech Communication and Technology (EuroSpeech '95), September 1995, Madrid, Spain* 783-786.

3. Young SJ, Brown MG, Foote JT, Jones GJF, Sparck Jones K: **Acoustic indexing for multimedia retrieval and browsing.** *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97), April 1997, Munich, Germany* **1:** 199-202.

4. Yu P, Seide F: **Fast two-stage vocabulary-independent search in spontaneous speech.** *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05), March 2005, Philadelphia, Pa, USA* **1:** 481-484.

IPR2022-01289 Page 00174

5. Thambirratnam K, Sridharan S: **Dynamic match phone-lattice search for very fast and accurate unrestricted vocabulary keyword spotting.** *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05), March 2005, Philadelphia, Pa, USA* **1:** 465-468.

6. Cheng YM, Ma C, Melnar L: **Voice-to-phoneme conversion algorithms for speaker-independent voice-tag applications in embedded platforms.** *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU '05), November-December 2005, San Juan, Puerto Rico* 403-408.

7. Liu C, Melnar L: **An automated linguistic knowledge-based cross-language transfer method for building acoustic models for a language without native training data.** *Proceedings of the 9th European Conference on Speech Communication and Technology (InterSpeech '05), September 2005, Lisbon, Portugal* 1365-1368.

8. Sooful JJ, Botha EC: **Comparison of acoustic distance measures for automatic cross-language phoneme mapping.** *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP '02), September 2002, Denver, Colo, USA* 521-524.

IPR2022-01289 Page 00175

9. Schultz T, Waibel A: **Fast bootstrapping of LVCSR systems with multilingual phoneme sets.** *Proceedings of the 5th European Conference on Speech Communication and Technology (EuroSpeech '97), September 1997, Rhodes, Greece* 371-373.

10. Schultz T, Waibel A: **Polyphone decision tree specialization for language adaptation.** *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00), June 2000, Istanbul, Turkey* **3:** 1707-1710.

11. Cheng YM, Ma C, Melnar L: **Cross-language evaluation of voice-to-phoneme conversions for voice-tag application in embedded platforms.** *Proceedings of the 9th International Conference on Spoken Language Processing (InterSpeech '06), September 2006, Pittsburgh, Pa, USA* 121-124.

12. Soong FK, Huang E-F: **A tree-trellis based fast search for finding the N-best sentence hypotheses in continuous speech recognition.** *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '91), May 1991, Toronto, Ontario, Canada* **1:** 705-708.

13. Hernandez-Abrego G, Olorenshaw L, Tato R, Schaaf T: **Dictionary refinements based on**

IPR2022-01289 Page 00176

phonetic consensus and non-uniform pronunciation reduction. *Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP '04), October 2004, Jeju Island, Korea* 1697-1700.

14. Cover TM, Thomas JA: *Element of Information Theory*. John Wiley & Sons, New York, NY, USA; 1991.

15. ES 201 108 : **Speech processing, transmission and quality aspects (STQ); distributed speech recognition; front-end feature extraction algorithm; compression algorithms.** 2000.

16. Melnar L, Talley J: **Phone merger specification for multilingual ASR: the Motorola polyphone network.** *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS '03), August 2003, Barcelona, Spain* 1337-1340.

17. Trask R: *A Dictionary of Phonetics and Phonology*. Routledge, London, UK; 1996.

Author information

Authors and Affiliations

Human Interaction Research, Motorola Labs, 1925 Algonquin Road, Schaumburg, IL,

IPR2022-01289 Page 00177

60196, USA

YanMing Cheng, Changxue Ma & Lynette Melnar

Corresponding author

Correspondence to Lynette Melnar.

Rights and permissions

Open Access This article is distributed under the terms of the Creative Commons Attribution 2.0 International License (https://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Reprints and Permissions

About this article

Cite this article

Cheng, Y., Ma, C. & Melnar, L. Voice-to-Phoneme Conversion Algorithms for Voice-Tag Applications in Embedded Platforms. *J AUDIO SPEECH MUSIC PROC.* **2008,** 568737 (2007). https://doi.org/10.1155/2008/568737

Received	Revised	Accepted
28 November 2006	15 July 2007	26 September 2007

Published
04 October 2007

DOI
https://doi.org/10.1155/2008/568737

Share this article

Anyone you share the following link with will be able to read

IPR2022-01289 Page 00178

this content:

Get shareable link

Provided by the Springer Nature SharedIt content-sharing initiative

Keywords

Speech Recognition **Target Language**

Dynamic Time Warping **Acoustic Model**

Source Language

Support and Contact Terms and conditions

Jobs Privacy statement

Language editing for authors Accessibility

Scientific editing for authors Cookies

Leave feedback

Follow SpringerOpen

By using this website, you agree to our Terms and Conditions, California Privacy Statement, Privacy statement and Cookies policy. Manage cookies/Do not sell my data we use in the preference centre.

SPRINGER NATURE

© 2022 BioMed Central Ltd unless otherwise stated. Part of Springer Nature.

IPR2022-01289 Page 00179

ATTACHMENT

V

LIBRARY | CATALOG

Search Navigation < 6 of 69 >

BOOK

Network+ study guide

Full Record MARC Tags

Main title

Network+ study guide / David Groth ... [et al.].

Edition

Deluxe ed.

Published/Created

San Francisco [Calif.] : Sybex, c2005.

Request this Item | LC Find It

Links

Links

Contributor biographical information http://www.loc.gov/catdir/enhancements/fy0628/2005920773-b.html
Publisher description http://www.loc.gov/catdir/enhancements/fy0628/2005920773-d.html
Table of contents only http://www.loc.gov/catdir/enhancements/fy0628/2005920773-t.html

More Information

LCCN Permalink https://lccn.loc.gov/2005920773

IPR2022-01289 Page 00181

Description	xxxviii, 519 p. : ill. ; 24 cm. + 2 CD-ROMs (4 3/4 in.)
ISBN	0782144160
LC classification	TK5105.5 .N46673 2005
Variant title	Network plus study guide
Related names	Groth, David.
LC Subjects	Computer networks--Examinations--Study guides. Telecommunications engineers--Certification--Study guides. Microsoft software--Examinations--Study guides.
Browse by shelf order	TK5105.5
Notes	"Second edition"--Cover. "Exam N10-003"--Cover. Includes index.
LCCN	2005920773
Type of material	Book

Item Availability

CALL NUMBER	TK5105.5 .N46673 2005 FT MEADE Copy 1
Request in	Machine Readable Collections - STORED OFFSITE
Status	c.1 Charged - Due - (Internal Loan)

CALL NUMBER	TK5105.5 .N46673 2005 Copy 2
Request in	See Reference Staff. By Appt in Jefferson Main RR (MRC)
Status	c.2 In Process 06-25-2007

IPR2022-01289 Page 00182

IPR2022-01289 Page 00183

ATTACHMENT

W

LIBRARY | CATALOG

1 of 1

BOOK

Network+ study guide

Full Record **MARC Tags**

000		01532cam a22003374a 4500
001		13844657
005		20070625122616.0
008		050118s2005 caua 001 0 eng
906	__	|**a** 7 |**b** cbc |**c** origres |**d** 2 |**e** epcn |**f** 20 |**g** y-gencatlg
925	0_	|**a** acquire |**b** 2 shelf copies |**x** policy default
955	__	|**a** pc17 2005-01-18 |**a** jf00 2006-03-06 |**i** jg12 2006-08-09 |**e** jg12 2006-08-09 |**a** ja15 2007-06-25 Copy 2 to BCCD
010	__	|**a** 2005920773
020	__	|**a** 0782144160
040	__	|**a** DLC |**c** DLC |**d** DLC
050	00	|**a** TK5105.5 |**b** .N46673 2005
245	00	|**a** Network+ study guide / |**c** David Groth ... [et al.].
246	3_	|**a** Network plus study guide
250	__	|**a** Deluxe ed.
260	__	|**a** San Francisco [Calif.] : |**b** Sybex, |**c** c2005.
300	__	|**a** xxxviii, 519 p. : |**b** ill. ; |**c** 24 cm. + |**e** 2 CD-ROMs (4 3/4 in.)
500	__	|**a** "Second edition"--Cover.
500	__	|**a** "Exam N10-003"--Cover.
500	__	|**a** Includes index.
650	_0	|**a** Computer networks |**x** Examinations |**v** Study guides.
650	_0	|**a** Telecommunications engineers |**x** Certification |**v** Study guides.
650	_0	|**a** Microsoft software |**x** Examinations |**v** Study guides.

IPR2022-01289 Page 00185

700	1_	|**a** Groth, David.
856	42	|**3** Contributor biographical information |**u** http://www.loc.gov/catdir/enhancements/fy0628/2005920773-b.html
856	42	|**3** Publisher description |**u** http://www.loc.gov/catdir/enhancements/fy0628/2005920773-d.html
856	41	|**3** Table of contents only |**u** http://www.loc.gov/catdir/enhancements/fy0628/2005920773-t.html
963	__	|**a** Dennis Fitzgerald; phone: (510) 523-8233; email: dfitzgerald@sybex.com; bc: dfitzgerald@sybex.com

Request this Item | LC Find It

Item Availability

CALL NUMBER	TK5105.5 .N46673 2005 FT MEADE Copy 1
Request in	Machine Readable Collections - STORED OFFSITE
Status	Not Charged

CALL NUMBER	TK5105.5 .N46673 2005 Copy 2
Request in	See Reference Staff. By Appt in Jefferson Main RR (MRC)
Status	c.2 In Process 06-25-2007

IPR2022-01289 Page 00186

ATTACHMENT

X

Alabama State University Library's HORNet Cat

Databases Search Headings Titles Patron Login Request History Help Exit

Database Name: HORNetCAT
Search Request: Builder = (57381883)[in]
Search Results: Displaying 1 of 1 entries

previous next

Brief Record | Detailed Record | Table of Contents | Linked Resources | MARC Format

Network+ study guide / David Groth ... [et al.].

000 01383cam a22003614a 450
001 260963
005 20070511124455.0
008 050118s2005 caua 001 0 eng
010 __ |**a** 2005920773
035 __ |**a** (OCoLC)ocm57381883
040 __ |**a** DLC |**c** DLC |**d** UKM |**d** IG# |**d** BAKER |**d** BTCTA
015 __ |**a** GBA501106 |**2** bnb
016 7_ |**a** 013074529 |**2** Uk
020 __ |**a** 0782144160
020 __ |**a** 9780782144161
035 __ |**a** (OCoLC)***57381883***
090 __ |**a** TK5105.5 |**b** .N47 2005
049 __ |**a** AMUA
245 00 |**a** Network+ study guide / |**c** David Groth ... [et al.].
246 3_ |**a** Network plus study guide
250 __ |**a** Deluxe ed.
260 __ |**a** San Francisco [Calif.] : |**b** Sybex, |**c** c2005.
300 __ |**a** xxxviii, 519 p. : |**b** ill. ; |**c** 24 cm. + |**e** 2 CD-ROMs (4 3/4 in.)
500 __ |**a** "Second edition"--Cover.
500 __ |**a** "Exam N10-003"--Cover.
500 __ |**a** Includes index.
650 _0 |**a** Computer networks |**x** Examinations |**v** Study guides.
650 _0 |**a** Telecommunications engineers |**x** Certification |**v** Study guides.

IPR2022-01289 Page 00188

650 _0 |**a** Microsoft software |**x** Examinations |**v** Study guides.

700 1_ |**a** Groth, David.

856 42 |**3** Contributor biographical information |**u** http://www.loc.gov/catdir/enhancements/fy0628/2005920773-b.html

856 42 |**3** Publisher description |**u** http://www.loc.gov/catdir/enhancements/fy0628/2005920773-d.html

856 41 |**3** Table of contents only |**u** http://www.loc.gov/catdir/enhancements/fy0628/2005920773-t.html

previous next

Record Options	
Select Download Format	Full Record Format for Print/Save
Enter your email address:	Email

Databases Search *Headings* Titles Patron Login Request History Help Exit

endeavor
contact us

ATTACHMENT

Y

Close this window to return to the catalogue

EXPLORE THE BRITISH LIBRARY

Item Details

FMT	BK
LDR	pam a2200229 a 45
001	013074529
003	Uk
005	20051102125559.0
008	041117s2005 caua 001 0 eng
010	|a 2005920773
015	|a GBA501106 |2 bnb
0167	|a 013074529 |2 Uk
020	|a 0782144160 : |c £59.99
040	|a StDuBDS |b eng |c StDuBDS |d Uk
042	|a ukblsr
08204	|a 004.6 |2 22
24500	|a Network+ : |b study guide.
250	|a [4th ed.] : Deluxe ed. / |b David Groth ... [et al.].
260	|a San Francisco, Calif. ; |a London : |b SYBEX, |c 2005.
300	|a xxxviii, 519 p. : |b ill ; |c 24 cm. + |e 2 CD-ROMs.
336	|a text |2 rdacontent
337	|a unmediated |2 rdamedia
338	|a volume |2 rdacarrier
500	|a Previous ed.: / David Groth, Todd Lammle and William Tedder. 2003.
500	|a Includes index.
500	|a Formerly CIP. |5 Uk
650 0	|a Computer networks |x Examinations |v Study guides.
650 0	|a Electronic data processing personnel |x Certification |v Study guides.
7001	|a Groth, David.
7001	|a Groth, David. |t Network+ study guide.
85241	|a British Library |b HMNTS |j EMC.2006.a.242
SYS	013074529

Accessibility Terms of use © The British Library Board

P2

ATTACHMENT

Z

GRETCHEN L. HOFFMAN

Professor

School of Library & Information Studies
Texas Woman's University
P.O. Box 425769
Denton, TX 76204
ghoffman@twu.edu

RESEARCH AND TEACHING INTERESTS

Library cataloging and classification, information organization, subject analysis.

EDUCATION

Ph. D., Library and Information Management 2008
Emporia State University, Emporia, Kansas
Dissertation: *Negotiating Normative Institutional Pressures and Maintaining Legitimacy in a Complex Work Environment: A Multiple Case Study of Three Academic Cataloging Units*

M.L.S. 2000
Emporia State University, Emporia, Kansas (Portland, Oregon cohort)

B.A. (High Honors). Major: Music; Minor: German 1995
University of Nevada, Reno, Reno, Nevada

EMPLOYMENT

Professor 2021 to present
School of Library & Information Studies
Texas Woman's University, Denton, Texas

Associate Professor 2014-2021
School of Library & Information Studies
Texas Woman's University, Denton, Texas

Assistant Professor 2008-2014
School of Library & Information Studies
Texas Woman's University, Denton, Texas

Lecturer II 2007-2008
School of Library & Information Studies
Texas Woman's University, Denton, Texas

Original Cataloger 2005-2007
University of Kansas Libraries, Lawrence, Kansas

Adjunct Instructor 2003-2005
School of Library and Information Management
Emporia State University, Emporia, Kansas

Enrollment Specialist 2003-2005
School of Library and Information Management
Emporia State University, Emporia, Kansas

Advising Assistant 2002-2003
School of Library and Information Management
Emporia State University, Emporia, Kansas

Documents Cataloging Librarian 2001-2002
Branford P. Millar Library
Portland State University, Portland, Oregon

Several paraprofessional librarian positions in serials acquisitions and cataloging 1995-2000

TEACHING EXPERIENCE

Courses Taught at Texas Woman's University (Graduate-level)

Cataloging and Classification
Advanced Cataloging and Classification
Collection Development
Indexing and Abstracting
Information Organization
Information Organization and Retrieval (no longer offered)
Technical Services (no longer offered)

Courses Taught at Emporia State University (Graduate-level)

Collection Development
Organizing Information
Theory of Organization of Information

Doctoral Dissertations

Enochs, Beth. 2016. "A Cautionary Tale: Mixed Methods Analysis of Elementary School Library Folklore Collections." Role: Committee member

Miller, Michelle. 2016. "Mature Learners and Information Literacy Instruction in Academic Libraries: How LIS Programs Prepare Academic Library Instructors for Adult Education Roles." Role: Committee member

McCourry, Maurine. 2014. "RDA and the Music Student: A Domain Analytic Approach to Catalog Needs." Role: Outside committee member

Sutton, Sarah. 2011. "Identifying Core Competencies for Electronic Resources Librarians in the 21st Century Library." Role: Committee member

Shelnutt, Cherri. 2011. "An Exploration of the Experiences of Pioneering Online Library and Information Science Graduate Students in the 1990s." Role: Committee member

Master's Theses

Enochs, Beth. 2009. "Rhyme or Reason? Identifying Distinguishing Features of Elementary School Library Poetry Collections." Role: Committee member

Advising and Mentoring

I advise approximately 100-120 master's degree students each semester.

PUBLICATIONS: BOOKS

Hoffman, Gretchen L., and Karen Snow, eds. 2021. *Cataloging and Classification: Back to Basics*. Abingdon, England: Routledge.

Hoffman, Gretchen L. 2019. *Organizing Library Collections: Theory and Practice*. Lanham, MD: Rowman & Littlefield.

-Won a 2020 Most Promising New Textbook Award (College) from the Textbook & Academic Authors Association.

PUBLICATIONS: PEER-REVIEWED ARTICLES AND BOOK CHAPTERS

Hoffman, Gretchen L., and Karen Snow, guest eds. 2021. "Cataloging and Classification: Back to Basics." Special issue, *Cataloging & Classification Quarterly* 59, no. 2-3.

-Lead guest editor of special issue published in two parts (11 articles).

Hoffman, Gretchen L., and Karen Snow. 2021. "Cataloging and Classification: Back to Basics: Introduction." *Cataloging & Classification Quarterly* 59, no. 2-3: 73-76.

Snow, Karen, Gretchen L. Hoffman, Maurine McCourry, and Heather Moulaison Sandy. 2018. "Phoenix or Dodo?: Re-Envisioning Cataloging Education." In Re-Envisioning the MLS: Perspectives on the Future of Library and Information Science Education, edited by Lindsay C. Sarin, Johnna Percell, Paul T. Jaeger, and John Carlo Bertot. *Advances in Librarianship*, 227-239. Bingley, UK: Emerald Publishing.

Hoffman, Gretchen L. 2017. "Subject Headings and Classification Literature, 2008-2011: A Bibliographic Essay." Chicago, Illinois: ALA CaMMS Research and Publications Committee. Available: http://alcts.libguides.com/c.php?g=699368&p=4963386

Snow, Karen, and Gretchen L. Hoffman. 2015. "What Makes an Effective Cataloging Course? A Study of the Factors That Promote Learning." *Library Resources & Technical Services* 59: 187-99.

Hoffman, Gretchen L. 2015. "What's the Difference between Soul Food and Southern Cooking? The Classification of Cookbooks in American Libraries." In *Dethroning the Deceitful Pork Chop: Rethinking African American Foodways from Slavery to Obama*, edited by Jennifer Jensen Wallach, 61-78. Fayetteville, AR: University of Arkansas Press.

-2017 Book Award, Association for the Study of Food and Society

Hoffman, Gretchen L. 2014. "Could the Functional Future of Bibliographic Control Change Cataloging Work? An Exploration Using Abbott." In *Functional Future for Bibliographic Control: Transitioning into New Communities of Practice and Awareness*, edited by Shawne D. Miksa, 59-74. London: Routledge.

Hoffman, Gretchen L. 2012. "Using the Quality Matters Rubric to Improve Online Cataloging Courses." *Cataloging & Classification Quarterly* 50: 158-71.

Hoffman, Gretchen L. 2012. "Could the Functional Future of Bibliographic Control Change Cataloging Work? An Exploration Using Abbott." *Journal of Library Metadata* 12: 111-26.

Hoffman, Gretchen L. 2010. "Negotiating Normative Institutional Pressures and Maintaining Legitimacy in a Complex Work Environment: A Multiple Case Study of Three Academic Cataloging Units." *Advances in Library Administration and Organization* 29: 243-92.

Hoffman, Gretchen L. 2010. "Response to Barbara Tillett, Re: Meeting Users' Needs in Cataloging: What is the Right Thing to Do?" *Cataloging & Classification Quarterly* 48: 346-47.

Hoffman, Gretchen L. 2009. "Meeting Users' Needs in Cataloging: What is the Right Thing to Do?" *Cataloging & Classification Quarterly* 47: 631-41.

Hoffman, Gretchen L. 2005. "The Statement of International Cataloguing Principles: A Global Cataloging Initiative." In *Libraries, Globalisation and Cooperation: Papers from the International Conference* held in Sofia, Bulgaria, 3-5 November 2004, edited by Herbert K. Achleitner and Alexander Dimchev, 187-192. Sofia, Bulgaria: "St. Kliment Ohridski," University of Sofia.

CONFERENCE PRESENTATIONS

Hoffman, Gretchen L. 2019. "The Needs of School and Public Library Students in Cataloging and Information Organization Courses." Presentation at the annual conference of the Association for Library and Information Science Education (ALISE), Denver, Colorado, September 24-26.

Hoffman, Gretchen L. 2019. "RDA Status Report and Update on the New Toolkit: Practice." Presentation at the annual conference of the Texas Library Association (TLA), Austin, Texas, April 15-18.

Snow, Karen, Gretchen L. Hoffman, Maurine McCourry, and Heather Moulaison Sandy. 2018. "Expanding the Universe of Cataloging Education: No Crystal Ball Required." Presentation at the annual conference of the Association for Library and Information Science Education (ALISE), Denver, Colorado, February 6-9.

Hoffman, Gretchen L. 2016. "Teaching Ethics in Technical Services: Developing Ethical Thinking and Decision-Making in Knowledge Organization Courses." Presentation at the annual conference of the Association for Library and Information Science Education (ALISE), Boston, Massachusetts, January 5-8.

Hoffman, Gretchen L. 2015. "Educating Catalogers: The State of the Art." Panel presentation at the American Library Association Midwinter Conference, Chicago, Illinois, January 30-February 3.

Snow, Karen, and Gretchen L. Hoffman. 2015. "Understanding the Elements That Help Students Learn Cataloging: A Research Study." Presentation at the annual conference of the Association for Library and Information Science Education (ALISE), Chicago, Illinois, January 27-30.

Snow, Karen, and Gretchen L. Hoffman. 2014. "What Makes an Effective Cataloging Course? A Study of the Factors That Promote Learning." Presentation at the American Library Association Annual Conference, Las Vegas, Nevada, June 26-July 1.

Hoffman, Gretchen L. 2013. "How Are Cookbooks Classified in Libraries? An Examination of LCSH and LCC." Paper presented at the fourth North American Symposium on Knowledge Organization, Milwaukee, Wisconsin, June 13-14. Proceedings available online: http://iskocus.org/nasko2013-proceedings.php

Hoffman, Gretchen L. 2013. "Online Cataloging Course Design: Some Suggestions." Presentation at the annual conference of the Association for Library and Information Science Education (ALISE), Seattle, Washington, January 22-25.

Hoffman, Gretchen L. 2009. "Applying the User-Centered Paradigm to Cataloging Standards in Theory and Practice: Problems and Prospects." Paper presented at the second North American Symposium on Knowledge Organization, Syracuse, New York, June 18-19. Proceedings available online: http://iskocus.org/nasko2009-proceedings.php

Hoffman, Gretchen L. 2009. "Meeting Users' Needs in Cataloging: What is the Right Thing to Do?" Paper presented at the first Ethics of Information Organization conference, Milwaukee, Wisconsin, May 22-23.

Hoffman, Gretchen L. 2005. "Controlling the Headings: A Case of Cataloging Policy Development at the Library of Congress." Poster presented at the doctoral student poster session of the annual conference of the Association for Library and Information Science Education (ALISE), Boston, Massachusetts, January 11-14.

Hoffman, Gretchen L. 2004. "The Statement of International Cataloguing Principles: A Global Cataloging Initiative." Paper presented at the international conference, Libraries, Globalisation and Cooperation, Sofia, Bulgaria, November 3-5.

WORKS IN PROGRESS

Hoffman, Gretchen L. "Standardization's Role in Organizing Library Collections: An Examination." *Cataloging & Classification Quarterly*.

HONORS AND AWARDS

Received

Most Promising New Textbook Award (College) from the Textbook & Academic Authors Association, 2020

Award for Outstanding Academic Mentor/Advisor, College of Professional Education, Texas Woman's University, 2018

Award for Outstanding Service, College of Professional Education, Texas Woman's University, 2017
Award for Distinction in Distance Education, Texas Woman's University, 2013

Award for Outstanding Teacher in Distance Education, College of Professional Education, Texas Woman's University, 2011

Beta Phi Mu: International Library & Information Studies Honor Society, Emporia State University, 2009

Nominated

Outstanding Faculty Member, Texas Woman's University, Graduate Student Council, Pioneering Spirit Awards, 2015

GRANTS AND FELLOWSHIPS

Texas Woman's University Travel Assistance Fund, 2019-2020, $800 grant to cover travel costs for presenting a paper at annual conference of the Association for Library and Information Science Education (ALISE), Knoxville, TN, 2019 (Funded)

Texas Woman's University Faculty Development Leave (Sabbatical), fall 2017

Texas Woman's University Travel Assistance Fund, 2014-2015, $800 grant to cover travel costs for presenting a paper at annual conference of the Association for Library and Information Science Education (ALISE), Chicago, IL, 2015 (Funded)

Texas Woman's University Faculty Development Grant, 2013-2014, $1,000, Principal Investigator (Funded)

Texas Woman's University Travel Assistance Fund, 2013-2014, $600 grant to cover travel costs for presentation at the annual conference of the American Library Association (ALA), 2014 (Funded)

Texas Woman's University Travel Assistance Fund, 2012-2013, $600 grant to cover travel costs for presentation at the annual conference of the Association for Library and Information Science Education (ALISE), 2013 (Funded)

Texas Woman's University, College of Professional Education Research Development Fund, 2011-2012, $1,300, Principal Investigator (Not funded)

American Library Association Carnegie-Whitney Grant, 2011-2012, $1,000, Principal Investigator (Not funded)

Texas Woman's University Travel Assistance Fund, 2008-2009, $600 grant to cover travel costs for presenting a paper at the North American Symposium on Knowledge Organization, Syracuse, NY, 2009 (Funded)

Texas Woman's University, College of Professional Education Research Development Fund, 2008-2009, $500, Principal Investigator (Funded)

Emporia State University, John and Valeta Richel Doctoral Fellowship, 2006, $3,000 (Funded)

PROFESSIONAL SERVICE

National Service

External Review Panel Member, American Library Association. Committee on Accreditation. Appointed April 2020

> Feb.-Mar. 2022: Served on an External Review Panel that reviewed a graduate program in library and information science (LIS)

Editorial Board Member of peer-reviewed journal *Library Resources & Technical Services*, American Library Association. Association for Library Collections & Technical Services (ALCTS). 2019-2021

Member, American Library Association. Association for Library Collections & Technical Services (ALCTS). Publications Committee, 2018-2020

Co-Chair, American Library Association. Association for Library Collections & Technical Services (ALCTS). Cataloging and Metadata Management Section (CaMMS). Research and Publications Committee, 2017-2019

Co-Chair, American Library Association. Association for Library Collections & Technical Services (ALCTS). Cataloging and Metadata Management Section (CaMMS). Competencies and Education for a Career in Cataloging Interest Group, 2018-2019

Co-Vice Chair, American Library Association. Association for Library Collections & Technical Services (ALCTS). Cataloging and Metadata Management Section (CaMMS). Competencies and Education for a Career in Cataloging Interest Group, 2017-2018

Peer Reviewer, Quality Matters, 2012-2018. I served on five review teams.

Convener, Technical Services Education Special Interest Group, Association for Library and Information Science Education (ALISE), 2015-2016

Member, Association for Library Collections & Technical Services (ALCTS) President's Symposium Planning Committee for the 2016 ALA Midwinter Conference, 2015-2016

Master Peer Reviewer, Quality Matters, 2012-2015. I chaired two review teams.

Discussion Leader, Cataloging & Metadata "Birds of a Feather" session, Association for Library and Information Science Education (ALISE) Annual Conference, 2013

Judge, Doctoral Poster Session, Association for Library and Information Science Education (ALISE) Annual Conference, 2011 and 2012

Co-Convener, Technical Services Education Special Interest Group, Association for Library and Information Science Education (ALISE), 2009-2011

Nominating Committee, Library Research Round Table, American Library Association, 2008-2009

Reviewer Activities

Reviewed two article submissions to *Cataloging & Classification Quarterly,* 2020-2021

Reviewed eight article submissions to *Library Resources & Technical Services*, 2019-2021

Reviewed book proposal about the *Dewey Decimal Classification* for Rowman & Littlefield, 2018

Reviewed book manuscript about cataloging ethics for McFarland, 2017

Reviewed book manuscript about *Library of Congress Classification*, for Rowman & Littlefield, 2017

Reviewed book proposal about the *Library of Congress Classification* for Rowman & Littlefield, 2016

Reviewed book proposal about the *Library of Congress Classification* for ABC-CLIO, 2016

Subject Matter Expert Reviewer for the publication *Competency Index for the Library Field.* 2014. Dublin, Ohio: OCLC WebJunction

Reviewed two article submissions to *Journal of Library Metadata*, 2012

Reviewed one article submission to *Cataloging & Classification Quarterly,* 2012

Reviewed five article proposals for a special issue of *Journal of Library Metadata*, 2011

Reviewed book manuscript about FRBR cataloging model for Neal-Schuman Publishers, 2009

State Service/Texas Library Community

Alternate Councilor, Cataloging and Metadata Round Table, Texas Library Association, 2013-2017

Past Chair, Cataloging and Metadata Round Table, Texas Library Association, 2012-2013

Chair, Cataloging and Metadata Round Table, Texas Library Association, 2011-2012

Chair-Elect, Cataloging and Metadata Round Table, Texas Library Association, 2010-2011

Texas Woman's University

Graduate Council, 2021 to present

Research Committee, 2021 to present

Faculty Senate, 2015-2017

Liaison to Distance Education Advisory Committee, 2016-2017
Budget and Planning Committee, 2015-2017
Academic Freedom Committee, 2015-2016

Distance Education Advisory Committee, 2016-2017

Strategic Planning Team, 2015-2016

University Curriculum Committee, 2015-2016

Teaching, Learning, and Curriculum (TLC) Master's degree program review committee, spring 2016

Mathematics-Teaching Master's degree program review committee, spring 2015

Reviewed grant proposal for TWU Research Enhancement Program, 2014

Quality Enhancement Plan (QEP) Advisory Council, 2013-2015

Faculty Development Leave Committee, 2013-2016

Reviewed grant proposals for TWU Quality Enhancement Plan (QEP) Project Grants, 2013

Student Code of Conduct Review Committee, 2012-2016

Library Advisory Committee, 2010-2014

SACS Subcommittee: Library & Other Learning Resources, 2010-2011

Traffic Committee, 2010-2012

TWU College of Professional Education

Strategic Planning Committee, 2014-2017

Goal 3 Leadership and Service Subcommittee, 2014-2015 (Chair)

Awards Committee, 2008 to fall 2014

TWU School of Library & Information Studies

Faculty Planning and Assessment Committee, 2009-2010; 2011-2013; 2014-2015; 2016-2017; 2018 to present; *Chair:* 2014-2015; 2020 to present

Peer Review Committee, 2015-2017; 2018-2021; *Chair:* 2018-2019; *Alternate:* spring 2022

Beta Phi Mu: International Library & Information Studies Honor Society, *Co-Advisor:* 2009-2010; *Advisor:* 2014-2017; 2018 to present

Enrollment Management Committee, 2009-2010, 2019-2020; *Chair*: spring 2010; 2019-2020

Curriculum Committee, 2007-2009; 2013-2014; 2016-2017; *Chair:* fall 2008; 2016-2017

Final Exam Committee, 2008-2009; 2015-2016; *Chair:* spring 2009

Information Organization and Retrieval (LS5013) Course Review Task Force, *Chair:* 2013-2016

Visioning Task Force, spring 2013-spring 2014

Library Science Federation Committee, 2008-2009; 2013-2014

Faculty Development Committee, 2010-2013; *Chair:* 2011-2013

Doctoral Committee, 2010-2011

Accreditation and Assessment Committee, 2008-2010; *Chair:* 2009-2010

Fall Forum Committee, 2008

PROFESSIONAL ORGANIZATIONS

American Library Association (ALA), 2003 to present

-Core: Leadership, Infrastructure, Futures
Formerly: Association for Library Collections & Technical Services (ALCTS)

-Library Research Round Table (LRRT)

Association for Information Science and Technology (ASIS&T), 2003 to present

-Classification Research Special Interest Group

Association for Library and Information Science Education (ALISE), 2003 to present

-Technical Services Education Special Interest Group

International Society for Knowledge Organization (ISKO), 2007 to present

Texas Library Association (TLA), 2007 to present

-Cataloging and Metadata Round Table
-College and University Libraries Division

Textbook & Academic Authors Association (TAA), 2019 to present

---