UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

MICROSOFT CORPORATION,
Petitioner,

v.

DIALECT, LLC,
Patent Owner.

IPR2025-01229
Patent: 7,634,409
Issued: December 15, 2009
Application No. 11/513,269
Filed: August 31, 2006

Title: DYNAMIC SPEECH SHARPENING

DECLARATION OF PAUL JACOBS

MICROSOFT CORP.
EXHIBIT 1003

TABLE OF CONTENTS

I, Paul Jacobs, do hereby declare as follows:

I. INTRODUCTION AND ENGAGEMENT

1. I have been retained as an independent expert on behalf of Microsoft Corporation in connection with the above-captioned Petition for *Inter Partes* Review ("IPR") to provide my analyses and opinions on certain technical issues related to U.S. Patent No. 7,634,409 (hereinafter "the '409 patent").

2. I am being compensated at my usual and customary rate for the time I spent in connection with this IPR. My compensation is not affected by the outcome of this IPR.

3. Specifically, I have been asked to provide my opinions regarding whether claims 1-3 and 6 (each a "Challenged Claim" and collectively the "Challenged Claims") of the '409 patent would have been obvious to a person having ordinary skill in the art ("POSITA") as of August 31, 2006. It is my opinion that each Challenged Claim would have been obvious to person of ordinary skill in the art after reviewing the prior art discussed herein.

II. BACKGROUND AND QUALIFICATIONS

4. My name is Paul Jacobs, and I am over 21 years of age and otherwise competent to make this Declaration. I make this Declaration based on facts and matters within my own knowledge and on information provided to me by others,

and, if called as a witness, I could and would competently testify to the matters set forth herein.

5. I have summarized in this section my educational background, career history, and other qualifications relevant to this matter. A more complete recitation of my professional experience including a list of my journal publications, patents, conference proceedings, book authorship, and committee memberships may be found in my curriculum vitae, which is attached hereto as Appendix A.

6. I am an expert in natural language processing and have a broad background in computer science that I have applied for decades as a scientist in the field of natural language processing, as a technology leader and executive, as a consultant, a professor, and as an advisor regarding software patents.

7. I received a Bachelor of Science in Applied Mathematics from Harvard University in 1981, a Master of Science in Applied Mathematics from Harvard in 1981, and a Ph.D. in Computer Science from the University of California at Berkeley in 1985.

8. I have authored or co-authored over 50 scientific and technical publications, primarily in the fields of artificial intelligence, natural language processing, and information retrieval. I am listed as an inventor on two U.S. patents directed to computational lexicons, and I have over 40 years of experience in the computer and information industry.

9. I have served in numerous professional and scientific capacities, including one year as a visiting professor of computer science at the University of Pennsylvania and several years as a member of the executive committee of the Association for Computational Linguistics. Currently, I serve on the Technology Policy Committee of the Association for Computing Machinery (USTPC). As an adjunct lecturer, I taught classes in the College of Information Studies (The "iSchool") at the University of Maryland in College Park from 2007. While my specialty has always been in artificial intelligence and natural language processing, teaching at the graduate level kept me current in a broad base of technologies pertaining to speech recognition and natural language processing systems.

10. Between 1985 and 1994, after completing my doctorate, I was employed as a computer scientist with General Electric ("GE") Corporate Research and Development, where I worked on a broad range of speech recognition and natural language processing research and development, including text processing and written and spoken language interfaces. I also consulted for Infonautics, an early Internet information services and advanced search company. I was the editor of a book, entitled "Text-Based Intelligent Systems." The book was a collection of papers based on a symposium I chaired in 1990, which brought together leaders of the field of Information Retrieval to address issues related to large-scale advanced text

processing. My work foreshadowed subsequent advances in AI that captured the power of large quantities of real data to simulate human-like capabilities.

11. During my years at GE, I was principal investigator for the GE team in the Tipster program, sponsored by the Advanced Research Projects Agency (ARPA) of the United States Department of Defense and other government agencies. The technology developed in Tipster formed the foundation of the first web search engines, and my work as an ARPA principal investigator in natural language kept me in close contact with research in related fields, particularly spoken language recognition. For example, I served on the program committee for ARPA's Speech and Natural Language Conference.

12. I joined SRA International ("SRA") in the latter part of 1994 and became director of media information technologies. My responsibilities included new ventures and technology related to the Internet and the World Wide Web. From 1994 until 2002, as the Web came of age, I held a series of technology and business management jobs in organizations focused on networked information management applications. For example, from late 1999 through the end of 2000, I was president and then CTO of AnswerLogic, one of the first companies directed toward answering naturally-expressed user questions on the Web. During this time we also experimentally developed spoken-language versions of our technology.

13. From my initial academic training in artificial intelligence, which was supervised by Marvin Minsky at MIT's Artificial Intelligence Laboratory while I was completing my Master's work at Harvard, through my career in commercial and academic R&D and as a graduate instructor, I was closely involved in the advances in speech and language technology. I visited Verbex, the predecessor of Dragon Systems, while at Harvard, made multiple invited trips to Bell Laboratories and while at Berkeley regularly participated in seminars and collaboration with SRI, where I witnessed the birth of the company that later became Nuance, to which the '409 patent was assigned (the commercial heritage of Nuance is confusing, because the name belonged to another company called Scansoft that was in a different business, but Nuance's speech technology that was later used in Siri came from the SRI spinoff that was purchased by Nuance). As a DARPA principal investigator, I participated regularly in some of the first Speech and Language Workshops (sometimes known as Human Language Technology workshops) where we collaborated and tested against shared "benchmarks". As mentioned above, I later collaborated with Nuance and others in integrating question answering technology with speaker-independent spoken language recognition.

14. Based on my experiences described above, and as indicated in my Curriculum Vitae, I am qualified to provide the following opinions with respect to the patents in this case. While my extensive background in these areas was well

beyond the capabilities of one of ordinary skill in August 2006 (and August 2005), I am qualified to opine on what one of ordinary skill would have known and understood as of the priority date of the '409 Patent.

15. A more complete recitation of my professional experience including a list of my journal publications, patents, conference proceedings, book authorship, and committee memberships may be found in my Curriculum Vitae, attached hereto as Appendix A.

III. MATERIALS CONSIDERED AND INFORMATION RELIED UPON REGARDING '409 PATENT

16. In preparing this declaration, I have reviewed the following materials bearing Exhibit Nos. that I understand are being referenced in the IPR petition which my declaration accompanies:

LIST OF EXHIBITS

No.	Description
1001	U.S. Patent No. 7,634,409 ("**'409 patent**")
1002	File History of U.S. Patent No. 7,634,409
1004	Provisional U.S. Patent Application No. 60/712,412
1005	PCT Patent Application Pub. No. WO 2007/027989 A2 to Di Cristo et al.
1006	EP Patent Application No. 06814053.2 to Robert A. Kennewick
1007	Claim Comparison Chart Of '409 Patent Claims 1-3 And 6 Against May 14, 2008, Amended Claims 9-11 And 14 of EP Application No. 06814053.2

No.	Description
1008	Bazzi, I., & Glass, J., *Heterogeneous Lexical Units For Automatic Speech Recognition: Preliminary Investigations*, Proc. of 2000 IEEE Int'l Conf. of Acoustics, Speech, and Signal Processing, 1257-1260 (2000) ("**Bazzi**")
1009	Huang, X., et al., *Spoken Language Processing – A Guide to Theory, Algorithm, and System Development*, Prentice-Hall, Inc. Publ. (2001) (excerpts) ("**Huang**")
1010	Jurafsky, D., & Martin, J., *Speech and Language Processing – An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice-Hall, Inc. Publ. (2000) (excerpts) ("**Jurafsky**")
1011	U.S. Patent No. 7,818,176 to Freeman et al. ("**'176 patent**")
1012	Claim Construction Order, *VB Assets, LLC v. Amazon.com, Inc.*, No. 19-1410 (MN) (D. Del. June 23, 2021)
1013	Institution Decision, IPR2020-01390, Paper 7 (PTAB Mar. 11, 2021)
1014	Livescu, K. et al., *Subword Modeling for Automatic Speech Recognition: Past, Present, and Emerging Approaches*, IEEE Signal Proc. Mag. 29:6, 44-57 (2012)
1015	Schmandt, C., *Voice Communications With Computers*, Int'l Thomson Publ. (excerpts) (1994)
1016	Ostendorf, M., & Roukos, S., *A Stochastic Segment Model For Phoneme-Based Continuous Speech Recognition*, IEEE Transactions On Acoustics, Speech, And Signal Proc., 1857-1869, 37:12 (Dec. 1989)
1017	Ravishankar, M., *Efficient Algorithms for Speech Recognition*, Thesis, CMU-CS-96-143, Carnegie Mellon Univ. (May 15, 1996)
1018	Rabiner, L., & Juang, B., *Fundamentals of Speech Recognition*, Prentice-Hall, Inc. Publ. (1993)

No.	Description
1019	U.S. Patent No. 6,085,160 to D'hoore et al.
1020	U.S. Patent No. 6,154,722 to Jerome R. Bellegarda
1021	U.S Patent No. 7,146,319 to Melvyn J. Hunt
1022	U.S. Patent Application Pub. No. 2004/0186714 to James K. Baker
1023	U.S. Patent No. 5,806,032 to Richard William Sproat
1024	U.S. Patent No. 6,108,627 to Michael Sabourin ("**Sabourin**")
1025	U.S. Patent Application Pub. No. 2005/0055209 A1 to Epstein et al. ("**Epstein**")
1026	Schwartz, R., & Chow, Y., *The N-Best Algorithm: An Efficient And Exact Procedure For Finding The N Most Likely Sentence Hypotheses*, IEEE Int'l Conf. of Acoustics, Speech, and Signal Processing, 81-84 (1990)
1027	U.S. Patent No. 5,241,619 to Richard M. Schwartz
1028	Mohri, M., et al., *Weighted finite-state transducers in speech recognition*, Computer Speech and Language, 16, 69-88 (2002)
1029	Declaration of Gordon MacPherson, signed June 23, 2025
1030	Proceedings, 2000 IEEE Int'l Conference on Acoustics, Speech, and Signal Processing (June 2020) (excerpts)
1031	Appendix of Challenged Claims 1-3 and 6

IV. UNDERSTANDING OF PATENT LAW

17. I am not an attorney. 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 was provided to me by the Petitioner's attorneys.

18. I understand that when considering the scope of the claims of a patent, the patent claim terms should generally be given the ordinary meaning that the terms would have to a person of ordinary skill in the art in question after reading the patent as of the earliest claimed priority date. As discussed below, it is my opinion that the challenged claims are not entitled to the earliest priority date asserted by the '409 patent. However, in my opinion, the meaning of the claim terms would not have materially differed to a person of ordinary skill in the art whether the claims were evaluated at the time of the earliest alleged priority date (August 31, 2005) or at the time the application that led to the '409 patent was filed (August 31, 2006).

19. I understand that the person of ordinary skill in the art is deemed to read the claim term not only in the context of the particular claim in which the term appears, but in the context of the entire patent, including the other claims and the specification of the application as filed. I further understand that the principal considerations regarding the scope and meaning of the claims are the plain language of the claim (including the surrounding claim language and context), the patent specification, and the prosecution history. I understand that while a claim is to be

read in light of the specification, one must generally avoid importing limitations into the claim from the specification. I am also informed that the prosecution history can often inform the meaning of the claim by demonstrating how the inventor understood the invention and whether the inventor limited the invention in the course of prosecution, making the claim scope narrower than it would otherwise be. I applied these understandings when considering the scope and meaning of the claims of the '409 patent.

20. I understand that a prior art reference anticipates an asserted claim, and thus renders the claim unpatentable, if all elements of the claim are disclosed in that prior art reference, either explicitly or inherently (i.e., necessarily present or implied).

21. I further understand that a claim is unpatentable if it would have been obvious. Obviousness of a claim requires that the claim would have been obvious from the perspective of one of ordinary skill in the art at the time the alleged invention was made. I understand that a claim could have been obvious from a single prior art reference or from a combination of two or more prior art references. I understand that an obviousness analysis requires an understanding of the scope and content of the prior art, any differences between the alleged invention and the prior art, and the level of ordinary skill in evaluating the pertinent art.

22. I further understand that a claim would have been obvious if it unites old elements with no change to their respective functions, or merely substitutes one element for another known in the field, and that combination yields predictable results. While it may be helpful to identify a reason for this combination, I understand that there is no strict requirement of finding an express teaching, suggestion, or motivation to combine within the references. When a product is available, design incentives and other market forces can prompt variations of it, either in the same field or different one. If one of ordinary skill in the art can implement a predictable variation, obviousness likely bars its patentability. For the same reason, if a technique has been used to improve one device and one of ordinary skill in the art would recognize that it would improve similar devices in the same way, using the technique would have been obvious. I understand that a claim would have been obvious if common sense directs one of ordinary skill, without hindsight bias caused by knowledge of the challenged claims, to combine multiple prior art references or add missing features to reproduce the alleged invention recited in the claims.

23. I further understand that certain factors may support or rebut the obviousness of a claim. I understand that such secondary considerations include, among other things, commercial success of the patented invention, skepticism of those having ordinary skill in the art at the time of invention, unexpected results of

the invention, any long-felt but unsolved need in the art that was satisfied by the alleged invention, the failure of others to make the alleged invention, praise of the alleged invention by those having ordinary skill or expertise in the art, and copying of the alleged invention by others in the field. I understand that there must be a nexus—a connection—between any such secondary considerations and the alleged invention. I also understand that contemporaneous and independent invention by others is a secondary consideration tending to show obviousness.

24. I am not aware of any allegations by the named inventors of the ’409 patent or any assignee of the ’409 patent that any secondary considerations tend to rebut the obviousness of any Challenged Claim of the ’409 patent.

25. I understand that in considering obviousness, it is important not to determine obviousness using the benefit of hindsight derived from the descriptions found in the patent being considered.

26. I understand that Petitioner has the burden of proving unpatentability by a preponderance of evidence, which means that the claims are more likely than not unpatentable.

27. The analysis in this declaration is in accordance with the above-stated legal principles.

V. THE '409 PATENT

28. The '409 patent, titled "Dynamic speech sharpening" issued on December 15, 2009. The '209 patent issued from U.S. Patent Application No. 11/513,269 (the "'269 application"), filed on August 31, 2006. While the '409 patent claims the benefit of provisional U.S. Patent Application Ser. No. 60/712,412 (the "'412 provisional", EX1004), filed Aug. 31, 2005, the '412 provisional does not support claim 1 of the '409 patent[1] and I understand that the challenged claims are therefore not entitled to a priory date earlier than their August 31, 2006, filing.

A. The '409 Patent's Specification

29. The '409 patent describes a system for interpreting natural language speech. The system is configured for "receiving a user verbalization," "identifying one or more phonemes in the verbalization," and using an "acoustic grammar … to map the phonemes to syllables" to generate "preliminary interpretations of the verbalization." EX1001, Abstract.

30. The '409 patent states that its acoustic grammar may represent the phonotactic rules of the English language. For example, the patent describes that for an "acoustic grammar representing the phonotactic rules of English" syllables may

[1] For example, the words "phonemes" or "syllables" are found nowhere in the provisional application. EX1004.

be "divided into core components of an onset, a nucleus, and a coda." *Id*., 6:21-26. The patent further describes that the phonotactic restraints for a given acoustic model may be used to restrict the "available transitions" between acoustic elements. *Id*., 7:26-29.

31. The system includes a module that generates multiple candidate interpretations of the phoneme sequence and assigns to each candidate a "confidence or interpretation score … representing a likelihood that a particular candidate interpretation is a correct interpretation of the verbalization." *Id*., 9:15-19. Based on the scores, the system selects the candidate with the highest or lowest value as a probable interpretation of the utterance. *Id*., 9:19-23.

B. The Prosecution History

1. Prosecution Of The '409 Patent

32. I have reviewed the file history of the '659 patent. I understand that the Examiner provided the following reasons for allowance, distinguishing the allowed claims over U.S. Patent No. 7,146,319 to Hunt:

> Hunt fails to specifically disclose mapping the recognized stream of phonemes to an acoustic grammar that phonemically represents one or more syllables, the recognized stream of phonemes mapped to a series of one or more of the phonemically represented syllables; and wherein the generated interpretation includes the series of

syllables mapped to the recognized stream of phonemes. **In other words, Hunt fails to teach matching phonemes against syllable grammars.**

EX1002, 148, 186-93 (emphasis added).

2. European Counterpart Prosecution

33. I understand that on the ’269 application’s filing date of August 31, 2006, the Applicant also filed corresponding PCT application no. PCT/US2006/034184 (EX1005), which shares the same drawings and written description as the ’269 application.

34. The applicant filed an initial amendment to its claims on May 14, 2008, before initial examination by the EPO. Id., 137-42. EX1007 presents an element-by-element comparison between the amended claims 9-11 and 14 and respective claims 1-3 and 6 of the ’409 Patent. As shown in EX1007, every limitation of the ’409 Patent’s method claims 1-3 and 6 has a substantially identical step in claims 9-11 and 14 of the EPO application as amended on May 14, 2008. EX1007. Following that initial amendment, the EPO filed a European search opinion on September 28, 2010, indicating that the examined claims were unpatentable as not novel based, in part, on the Bazzi reference (EX1008) relied on in this Petition. (EX1006, 100-04). Specifically, the European search opinion determined that Bazzi disclosed the following:

- A system for providing out-of-vocabulary interpretation capabilities and for tolerating noise when interpreting natural language speech utterances;
- at least one input device that receives an utterance from a user and generates an electronic signal corresponding to the utterance;
- a speech interpretation engine that receives the electronic signal corresponding to the utterance operable to:
    - recognize a stream of phonemes contained in the utterance;
    - map the recognized stream of phonemes to an acoustic grammar that phonemically represents one or more syllables, the recognized stream of phonemes mapped to a series of one or more of the phonemically represented syllables; and
    - generate at least one interpretation of the utterance, wherein the generated interpretation includes the series of syllables mapped to the recognized stream of phonemes.

Id., 102-03.

35. I understand that the European counterpart application did not result in an issued patent and was abandoned, and the application was closed on July 3, 2018. *Id*., 1-2.

VI. LEVEL OF SKILL IN THE ART AND STATE OF THE ART

A. Person Of Ordinary Skill In The Art

36. The person of ordinary skill in the art in August 2006 ("POSITA") would have had a bachelor's degree in electrical engineering, computer science, computer engineering, or equivalent field, and two years of experience working with speech recognition and natural language processing systems. Additional work experience could make up for less education and vice versa. This definition would not differ meaningfully were the date in question August 2005. In this Declaration, I use the terms "one of ordinary skill" and "skilled artisan" as synonymous with the "person of ordinary skill in the art."

B. State Of The Art

37. The below sub-sections describe concepts and processes that reflect the state of the art for speech recognition systems during the 2005-2006 time period as described in as described in textbooks such as Huang, X., et al., *Spoken Language Processing – A Guide to Theory, Algorithm, and System Development* (Prentice Hall Publ., 2001) ("Huang", EX1009), and Jurafsky, D., & Martin, J., *Speech and Language Processing – An Introduction to Natural Language Processing, Computation Linguistics, and Speech Recognition* (Prentice Hall Publ., 2000) ("Jurafsky", EX1010).

1. Linguistic Units In Speech Recognition

38. Automatic Speech Recognition (ASR) systems aim to identify the most likely sequence of words corresponding to a given speech input, accounting for uncertainties in a received signal such as pronunciation variability, ambient noise, and spontaneous speech disfluencies (e.g., spoken "uhs" and "ums," stutters, and word repetitions). *See* EX1010, 194-95; EX1009, xxii, 53. "Speech is based on a sequence of discrete sound segments that are linked in time. These segments, called phonemes, are assumed to have unique articulatory and acoustic characteristics…. Each phoneme has distinguishable acoustic characteristics and, in combination with other phonemes, forms larger units such as syllables and words." EX1009, xxii.

a) Phones and Phonemes Were Known As Interchangeable Acoustic Elements In Speech Recognition Systems

39. The basic acoustic-linguistic segment in speech recognition is the phoneme. *Id*., 37. ("In speech science, the term *phoneme* is used to denote any of the minimal units of speech sound in a language that can serve to distinguish one word from another."). A phoneme is either a consonant or a vowel, with the English language containing 16 vowel and 24 consonant phonemes. *Id*., xxii. When spoken, a phoneme may be articulated differently based on its surrounding phonemes; this is called "coarticulation." *Id*., 48. "[W]hen the variations resulting from coarticulatory processes can be consciously perceived, the modified phonemes are called

allophones." *Id*, 48. A phoneme may be articulated differently based on its surrounding phonemes, and the spoken articulation of a given phoneme is called a *phone*. Thus, a *phoneme* is a conceptual representation of a given speech sound providing potential meaning distinctions for a given language (i.e., how a sound is reflected lexically), while a *phone* constitutes a spoken articulation of a phoneme. [2] *Id.*, 37 ("We conventionally use the term *phone* to denote a phoneme's acoustic realization."); *see also* EX1010, 104.

40. As a *phone* is just an articulation of a given *phoneme*, persons of ordinary skill in the art often used and understood the terms "phoneme" and "phone" interchangeably, as reflected in the prior art literature. *See* EX1015, 15, ("For most practical purposes, phone and phoneme may be considered to be synonyms."); EX1009, 37 ("We will use the terms phone or phoneme interchangeably to refer to

[2] As an example, English speakers generally express the phoneme /p/ differently in pronouncing the words "pin" and "spin," aspirating the "p" sound in the former but not the latter. EX1009, 48. In other words, each of the words "pin" and "spin" realizes a different allophone of the phoneme /p/. In speech science, phonemes are traditionally written inside slashes, whereas allophonic variations of a given phoneme are written in brackets. For example, "/t/ is a phoneme whose allophones include [t^h], [ɾ], and [t̪]." EX1010, 104.

the speaker-independent and context-independent units of meaningful sound contrast."); EX1014, 45 ("In speech recognition research, these terms [phone and phoneme] are often used interchangeably, and recognition dictionaries often include a mix of phones and phonemes."); EX1016, 1858 ("In this paper, we are not rigorous in distinguishing between the two terms phone and phoneme, which are used interchangeably."); EX1017, 3 ("The *phoneme* (or phone) has been the most commonly accepted sub-word unit.")

b) Syllables (Each Having An Onset, Nucleus, And Coda) Were Known And Used Widely In Speech Recognition

41. To further constrain recognition and improve interpretability, speech recognition systems may segment phoneme/phone streams into syllables. Syllables are intermediate sub-word units "that interpose between the phones and the word level." EX1009, 51. A syllable consists of a vowel and its surrounding consonants; these building blocks of a syllable are called the onset, nucleus, and coda. See, e.g., EX1010, 102 ("A syllable is usually described as having an optional initial consonant or set of consonants called the onset, followed by a vowel or vowels, followed by a final consonant or sequence of consonants called the coda."). Huang provides an example segmenting the syllable "strengths" into the phonemes corresponding to its onset, nucleus, and coda components:

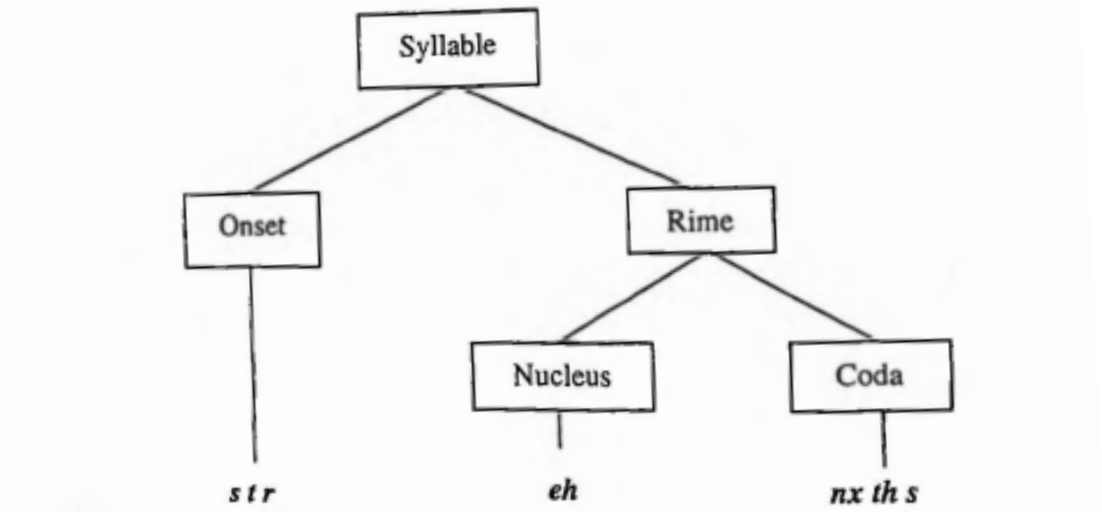


Figure 2.25 The word/syllable *strengths* (*/s t r eh nx th s/*) is the longest syllable of English.

EX1009, 52, Fig. 2.25; *see also* EX1001, 2:50-52 ("Portions of a word may be represented by a syllable, which may be further broken down into core components of an onset, a nucleus, and a coda."). As shown in Haung's Figure 2.25, for the syllable "strengths," the phonemes "s," "t," and "r" represent the syllable's onset (i.e., the initial consonants), the phoneme "eh" represents the nucleus (i.e., the vowel), and the phonemes "nx," "th," and "s" represent the coda (i.e., the final consonants). Huang recognizes that "syllables are often used" as subword models for large-vocabulary speech recognition systems. EX1009, 608; *see also* EX1018, 436-37 (describing using syllables as one of "several possible choices for subword units that can be used to model speech").

2. Automatic Speech Recognition Systems

42. Spoken language processing encompasses a range of computational techniques, algorithms, and statistical models used to convert human speech into structured representations such as word hypotheses. As described in Huang, a typical speech recognizer comprises a computing platform including components for input signal processing and a decoder module driven by acoustic and language models, as shown in Huang's Figure 1.2, reproduced below. EX1009, 5, Fig. 1.2.

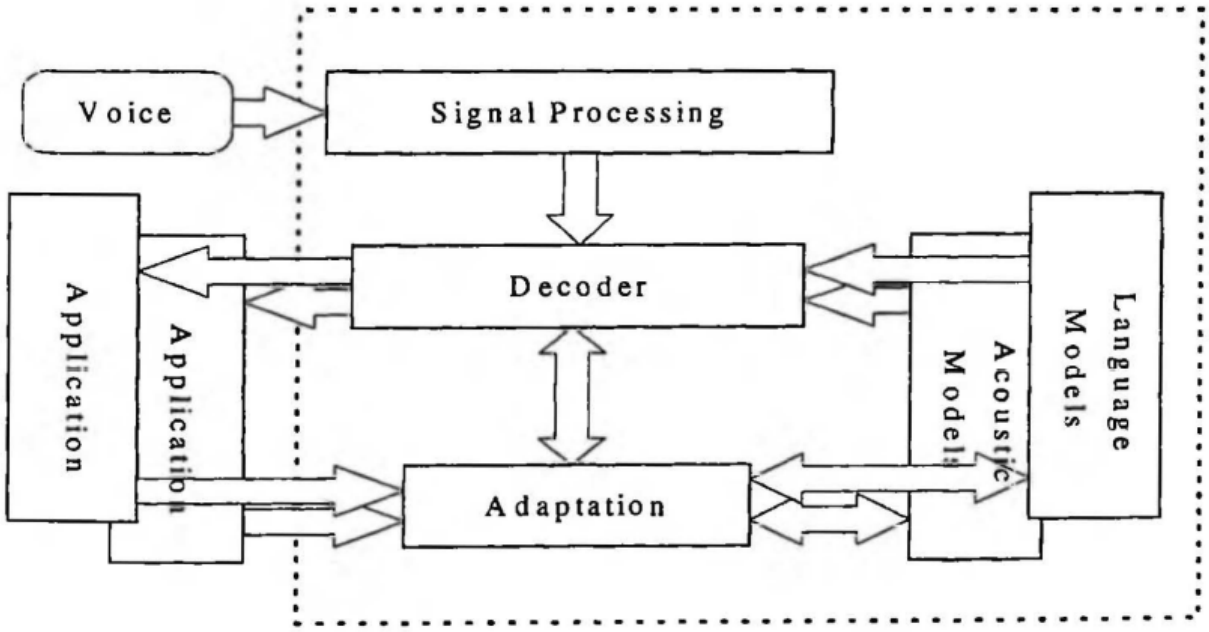


Figure 1.2 Basic system architecture of a speech recognition system [12].

43. ASR systems are typically structured around a processing pipeline involving steps of signal processing, phonetic analysis, and decoding. *See* EX1010, 240-41, Fig. 7.2. Jurafsky depicts these functional stages of a typical ASR system in Figure 7.2, reproduced below.

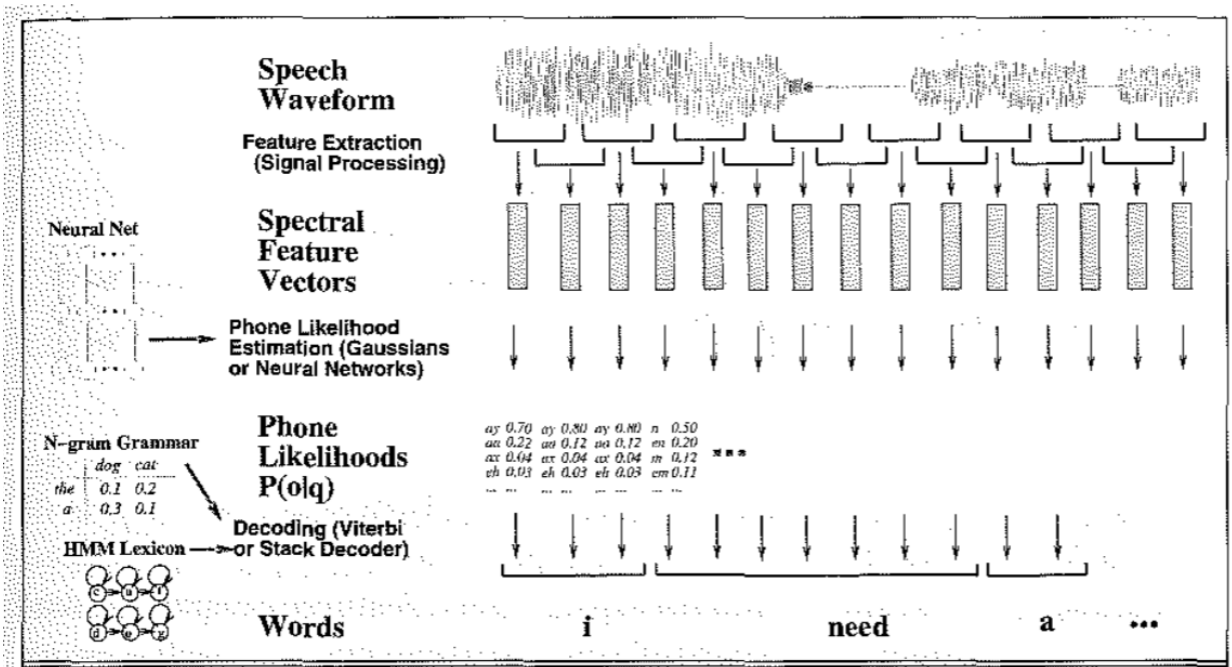


Figure 7.2 Schematic architecture for a (simplified) speech recognizer.

Id., 241. These functional stages are described in more detail below.

Signal Processing – In the first stage, signal processing, the acoustic signal is received as a waveform "which is transformed into spectral features which give information about how much energy in the signal is at different frequencies. *Id*., 240

Subword Recognition – In the second stage, subword or phone recognition, statistical techniques are used "to tentatively recognize individual speech sounds" and thereby identify the probabilities of what subword units are present in each time frame of the input signal. *Id*.

Decoding – In the final stage, decoding, the ASR system combines the sub-word probability data determined in the previous stage with "a dictionary of word pronunciations and a language model (probabilistic grammar)" using a decoding

algorithm "to find the sequence of words which has the highest probability given the acoustic events." *Id*., 241

3. Lexicons And Probabilistic Grammars Were Well Known In ASRs

44. As explained in the preceding sub-section, typical ASR systems identify a sequence of words corresponding to an input utterance by analyzing probable sub-word units in combination with two data structures: 1) a pronunciation dictionary (also called a lexicon) and 2) a probabilistic grammar (also called a language model). *See* EX1010, 270 (describing how a typical ASR decoder "takes 3 inputs (the observation likelihoods, the HMM lexicon, and the *N*-gram language model) and produces the most probable string of words").

45. A pronunciation dictionary maps words to their pronunciations as a set of phones. *Id*., 135 ("[Pronunciation dictionaries] give the pronunciation of words as strings of phones, sometimes including syllabification and stress."). The most common data structure type for representing a pronunciation dictionary is the Hidden Markov Model (HMM), which is built "by taking an off-the-shelf pronunciation dictionary" and mapping each phone/phoneme in the dictionary to a state in the HMM data structure. *Id*. 272.

46. The probabilistic grammar, meanwhile, models the phonotactic constraints of a language. That is, sub-word units such as phonemes and syllables in

natural languages are not freely combinable; languages impose phonotactic rules that restrict the permissible sequences of sub-word units. Likewise, given words in a language are more or less likely to follow other words. *See* EX1010, 191-92. ("Guessing the next word (or word prediction) is an essential subtask of speech recognition…. [L]ooking at previous words can give us an important cue about what the next ones are going to be."). Language models are commonly modeled using a data structure called an n-gram. "An N-gram model uses the previous N-1 words to predict the next one." *Id*., 193. For example, a word "bigram" models the probability of two words being in sequence while a "trigram" models the probability for three sequential words. *Id*., 197-98. Such language models help reduce ambiguity, provide tolerance for word misrecognitions caused by noise, and improve recognition accuracy by favoring more probable phoneme sequences and by pruning likely invalid phoneme sequences from the search space.

4. Finite State Transducers ("FSTs") Were Well Known

47. A finite-state transducer (FST) is a mathematical model used to represent mappings between sequences in the form of input-output symbol pairs, where each mapping may be associated with a cost or weight (called a weighted FST). EX1028, 69. "Therefore, a path through the transducer encodes a mapping from an input symbol sequence to an output symbol sequence." *Id*. FST's and their use in spoken language recognition were well known in the art prior to the 2000s.

Id., 70. In the context of speech recognition, FSTs are used to model and combine various components of the recognition process in a unified, efficient framework. FSTs allow for probabilistic transitions from one state to another and support "composition," i.e., the combination of probabilities and constraints among multiple FSTs. *Id*., 69-70, 73. FSTs are therefore widely employed to encode and combine probabilistic models such as pronunciation lexicons, language models, and observed acoustic feature probabilities. *Id*. These components are represented as individual FSTs and then composed together into a single search graph, which is searched using an appropriate search algorithm. *See* EX1008, 1258. This integrated transducer enables the system to efficiently evaluate possible interpretations of a speech input and identify the most likely hypothesis.

VII. CLAIM CONSTRUCTION

A. Claim 1 Preamble

48. I understand the preamble of claim 1 to be non-limiting. The preamble of claim 1 recites the following: "A method for providing out-of-vocabulary interpretation capabilities and for tolerating noise when interpreting natural language speech utterances." The preamble of claim 1 states the intended purpose or result of the claimed method—namely, to "provid[e] out-of-vocabulary interpretation capabilities" and to "tolerate noise" in interpreting speech. I understand that,

generally, a preamble is not limiting unless it recites essential structures or steps, or is necessary to give life, meaning, and vitality to the claim.

49. Here, nothing in the body of claim 1 references the goals of tolerating noise or handling out-of-vocabulary input. The claimed method begins with "receiving an utterance" and proceeds through steps involving phoneme recognition and syllabic mapping. These steps are described independently of any functional result like error tolerance or vocabulary adaptation. Moreover, the specification discusses those results as motivating factors behind the design, not as limitations on the method itself. Specifically, the patent's Background of the Invention describes purported problems in existing speech recognition systems relying on word-based grammars:

> In addition to the performance problems associated with speech recognition engines that employ large word grammars, existing speech processing engines are unable to interpret natural human speech with a suitable accuracy to sufficiently control some electronic devices. In particular, **speech interpretation engines still have substantial problems with accuracy and interpreting words that are not defined in a predetermined vocabulary or grammar context. Poor quality microphones, extraneous noises, unclear or grammatically incorrect speech by the user, or an accent of the user may also cause shortcomings in**

accuracy, such as when a particular sound cannot be mapped to a word in the grammar.

EX1001, 1:65-2:9. The specification proceeds to explain that out-of-vocabulary interpretation and noise reduction is merely an intended benefit that "may" be provided by phoneme recognition. *Id*., 2:40-44 ("Phoneme recognition may disregard the notion of words, instead interpreting a verbalization as a series of phonemes, which may provide out-of-vocabulary (OOV) capabilities, such as when a user misspeaks or an electronic capture devices [sic] drops part of a speech signal."; s*ee also id.*, 6:6-9 ("Phonemic recognition provides several benefits, particularly in the embedded space, such as offering out-of-vocabulary (OOV) capabilities."). The specification provides no other descriptions of out-of-vocabulary or noise reduction capabilities. Nor does it explicitly identify any particular steps for providing such capabilities. Moreover, the preamble provides no antecedent basis for terms appearing in the body of claim 1 or any dependent claims.

B. <u>"acoustic grammar"</u>

50. Claims 1, 2, and 3 of the '409 patent recite the term "acoustic grammar." In this Declaration, I give the term "acoustic grammar" its plain and ordinary meaning. However, I understand that the '409 patent shares the term

"acoustic grammar" with a non-family member patent[3] that has been construed in litigation before the District Court for the district of Delaware (the "Amazon Litigation") and by the PTAB in IPR2020-01390 (the "Amazon IPR"). These constructions are presented in the table below:

Term	Claim Construction Order[4] in D. Del.	Institution Decision[5] (Paper 7) in IPR2020-01390.
Acoustic grammar	"grammar of phonotactic rules of the English language that maps phonemes to syllables"	"collection of the phonemes, or distinct units of sound of a spoken language, linked together to form syllables, which are linked together to form the words of the language"

51. To the extent the Board determines that either of the constructions of "acoustic grammar" as interpreted in the Amazon Litigation or Amazon IPR are appropriate, this Declaration explains why the presented prior art also satisfies the language of the claims under such constructions.

[3] U.S. Patent No. 7,818,176 (the "'176 patent").

[4] EX1012.

[5] EX1013 at 12.

VIII. GROUND 1: CLAIMS 1, 2, 3, AND 6 ARE OBVIOUS OVER BAZZI

52. As explained below, it is my opinion that claims 1, 2, 3, and 6 are obvious over Bazzi.

A. Bazzi

53. Bazzi is a paper titled "HETEROGENEOUS LEXICAL UNITS FOR AUTOMATIC SPEECH RECOGNITION: PRELIMINARY INVESTIGATIONS" authored by Issam Bazzi and James Glass of the Spoken Language Systems Group at the Massachusetts Institute of Technology. EX1008, 1257. Bazzi was presented and distributed at the Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing and was subsequently also publicly available through IEEE's digital library as of August 6, 2002. *E.g.*, EX1029 (IEEE declaration describing distribution and publication of Bazzi); EX1030 (cover page and table of contents excerpt of the Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing showing inclusion of Bazzi, IEEE insignia, 2000 copyright, and ISBN and ISSN numbers for published proceedings). The EPO's reliance on Bazzi in the European counterpart prosecution is further evidence of Bazzi's public availability. I understand that Bazzi is therefore prior art under at least 35 U.S.C. § 102(b).

54. Bazzi describes a speech recognition system with a two-stage recognizer. As opposed to conventional single-stage recognition systems, such as

that described in the State of the Art Section (Section VI.B., above), which determine probable phones from a speech signal and convert those directly to word hypotheses, Bazzi describes a recognizer where, based on a graph of phonetic probabilities, "syllable graphs are computed in the first stage, and passed to the second stage to determine the most likely word hypotheses." EX1008, 1257-58.[6] Bazzi builds on the SUMMIT segment-based speech recognition system developed by MIT's Spoken Language Systems Group. *Id*., 1258(1:25-28). Bazzi calls its two-stage recognizer a "Syllable Recognizer" (Section 2.2) (EX1008, 1258(2:22)-1259(1:33)).

1. Claim 1

a) Element [1.1] [7]

[1.1] A method for providing out-of-vocabulary interpretation capabilities and for tolerating noise when interpreting natural language speech utterances, the method comprising:

[6] In addition to its syllable recognizer, Bazzi also proposes an alternative two-stage phone recognizer. *Id*., 1258(1:38-2:21). Bazzi's experimental testing, however, determined the syllable-based recognizer to be superior to the phone-based recognizer in terms of tested word error rate in word recognition. *Id*., 1259-60.

[7] Reference numbers in the format of [claim#.limitation#] are added throughout for ease of reference.

55. As explained in Section VII.A., above, I understand that the preamble of the ’409 patent is non-limiting.

56. To the extent the preamble is limiting, Bazzi renders it at least obvious. Bazzi discloses natural language speech processing “methods that can be used to model out-of-vocabulary and partial words.” EX1008, 1257(2:21-22). It provides these methods to address problems posed by the phenomena of “out-of-vocabulary words” and “partially spoken words, which are typically produced in more conversational or spontaneous speech applications.”[8] *Id*., 1257(2:17-19).

57. Bazzi’s methods, which “can be used to model out-of-vocabulary … words” use “more flexible sub-word units (such as phones or syllables, which are not constrained to match the active word vocabulary).” *Id.*, 1257(2:22-24). Bazzi teaches using these methods both “within a domain-dependent word-based recognition architecture” and as “a separate first stage, operating independently of a given vocabulary.” *Id.,* 1257(2:26-28). The ’409 patent’s specification does not explain how its methods specifically perform out-of-vocabulary recognition, instead treating such functionality to be a result of phoneme recognition. EX1001, 2:40-46. (“Phoneme recognition may disregard the notion of words, instead interpreting a

[8] Years later, the ’409 patent noted similar problems. EX1001, 1:65-2:9.

verbalization as a series of phonemes, which may provide out-of-vocabulary (OOV) capabilities.").

58. Relative to the '409 specification, Bazzi provides more extensive explanation of why recognition at the subword level (such as using phones and syllables) improves out of vocabulary recognition. Bazzi notes that reliance on a word-only lexicon can lead to erroneous interpretations of conversational speech. EX1008, 1257(2:18-20) ("These phenomena also tend to produce errors since the recognizer matches the phonetic sequence with the best fitting words in its active vocabulary."). The reference elaborates by noting that both phones and syllables have the "attractive property of being a "closed set, and thus will be able to cover new words." *Id.*, 1257(2:23-25). In other words, it is not practical for a recognition system to include all of the words that may be spoken, but it is feasible to include all of the possible phones, or all of the possible syllables. With respect to syllables in particular, Bazzi further establishes that a manageable syllable vocabulary of 1000 syllables can cover a word vocabulary of "around 45,000 words" (*Id*., 1259 and Figure 1). A skilled artisan would have understood that one of the goals of Bazzi's syllable-based recognizer, while not fully explored, was to provide superior out-of-vocabulary capabilities, and in fact Bazzi's syllable-based recognizer surely would have done so: The syllable lexicon was constructed from JUPITER's [word] vocabulary of 1957 words (*Id*.), yet it contained 1624 syllables, which based on

Bazzi's explanation would have been able to cover a vocabulary of closer to 50,000 words. Accordingly, one of ordinary skill would recognize Bazzi's disclosure of sub-word (including phones and syllables) based recognition to teach out-of-vocabulary recognition capabilities to at least the same extent the '409 patent does.

59. In addition to providing out-of-vocabulary interpretation capabilities, one of ordinary skill would have understood that Bazzi's more flexible speech recognition methods tolerate noise at least because they would better process partial word inputs caused by a noisy environment (just as they process partial words caused by "partial word utterances"). The '409 patent itself provides no meaningful discussion of "noise" much less an explanation of how its methods tolerate noise. Moreover, any speech recognition method (including those described by Bazzi) would be expected to tolerate some level of noise. *See e.g.*, EX1010, 145 (describing the "noisy channel model" for speech recognition that assumes that noise is introduced between a word's source and its receipt by a speech recognition system). Bazzi, on the other hand recognizes that its use sub-word units can cover partial word utterances (EX1008, 1257(2:23-25)); partial word utterances include situations where only a portion of spoken utterance is received by a recognizer due to noise. Most speech recognition systems during the 1980s and 1990s were tested on "benchmarks" that involved databases of continuous speech from multiple speakers in noisy environments, such as telephone "switchboards" and flight control systems.

Bazzi's experiments used the JUPITER system, which collected data from telephone handsets of people "dialing in" for information about the weather. At the time those data were collected, telephones were known to be noisy. In addition, Bazzi utilizes bigram and trigram language models, which provide noise tolerance by predicting potentially misrecognized syllables and words based on recognized preceding syllables and words. *See* EX1008, 1258(2:32-33).

60. Thus, Bazzi teaches a method that provides out-of-vocabulary interpretation capabilities and tolerate noise when interpreting natural language speech utterances.[9] As explained in the following sub-sections, Bazzi's methods at least render obvious each element of claim 1.

b) Element [1.2]

[1.2] receiving an utterance from a user;

[9] While Bazzi's "initial investigation" did not model or examine system behavior on out-of-vocabulary or partial words (EX1008, 1258(1:15-17)) Bazzi expressly states that its methods are intended to address out-of-vocabulary and partial word phenomena (*id.*, 1257-58) and that "preliminary results are quite encouraging" (*id.*, 1260(2:8-9)).

61. Bazzi discloses that its two-stage speech recognizer receives spoken utterances from users. For example, it describes a "two-stage recognizer configuration" in which "**a user** interacting with several different spoken dialogue domains (e.g., weather, travel, entertainment), **might have their speech initially processed by a domain-independent first stage, and then subsequently processed by domain dependent recognizers**." *Id.,* 1257(2:39)-1258(2:8) (emphasis added); *see also id.* (describing handling of "partial word utterances"). This is consistent with one of ordinary skill's understanding of the art that speech recognition systems are fundamentally designed and intended to receive a user's speech and process that speech into recognized language units. *See* EX1009, 5 (depicting and describing the "Basic system architecture of a speech recognition system," including starting with receiving a user's voice into a signal processing module). Throughout, Bazzi describes its test implementation of a "weather information system" that receives words spoken by a user. *E.g.*, EX1008, 1257(1:34-35) ("in our weather information system, we are constantly faced with new words spoken by users"); *id.*, 1258(1:16-18) (describing recognizing "within-vocabulary utterances"); *id.*, 1259(2:19) (describing a "training set" of utterances for the recognizer); *id.* 1260(2:23-24) (describing experimental results that varied "depending on the length of the utterance"). Accordingly, Bazzi discloses receiving speech (i.e., an utterance) from a user. At the very least, Bazzi renders receiving an

utterance obvious because its speech recognizer is premised on receiving spoken utterances from users, and then ultimately recognizing words based on those received spoken utterances.

c) Element [1.3]

[1.3] recognizing a stream of phonemes contained in the utterance on an electronic device;

62. Bazzi teaches recognizing a stream of phonemes contained within a given user utterance in that Bazzi teaches its Syllable Recognizer traversing a "scored phonetic graph" to derive a series of phonemes contained in the utterance. With respect to its Syllable Recognizer, Bazzi teaches creating a scored phonetic graph *P* that includes phones (i.e., phonemes) recognized in the received user utterance. Bazzi represents its two-stage syllable-based recognizer as an FST with the following formula, where *S* constitutes the search space (i.e., all possible paths in the FST graph) generated by composing various component FSTs and the operator "∘" denotes composition of component FSTs.

$$S = P \circ L_s \circ G_s \circ L_w \circ G \qquad (4)$$

EX1008, 1258(2:25). Bazzi calls the first composed FST, *P*, the "scored phonetic graph." *Id*., 1258(1:31) A scored phonetic graph constitutes a weighted graph of probable "phonetic units" Id., 1258(2:23-29) corresponding to the phonetic representation of the received acoustic signal. As explained in for the State of the

Art (Section VI.B.2., above), such a step of subword recognition to identify probabilities of subword units was well-known in the art. S*ee also* EX1010, 240-41, Fig. 7.2 (disclosing a "Phone Likelihoods" graph structure for conventional frame-based recognizer system). In the context of Bazzi's description of the FSTs in Equation (4), the scored phonetic graph to be a transducer that derives ("transduces") a series of phonemes, the graph itself representing a set of states with arcs indicating the possible subsequent phones at each state. By "scoring" the probable phonetic units contained in the received signal, probabilities at each transition state can be further constrained by (and combined with) relevant lexicons and grammar. *See* EX1010, 240-41 (combining scored phone likelihood graph with grammar and lexicon). The phonetic units derived in the scored phonetic graph P, are subsequently composed with a syllable lexicon, designated as "$L_{s..}$" EX1008, 1258(2:23-26). The syllable lexicon represents a mapping of phonetic units to syllables created from the relevant word lexicon by "partition[ing] the phone sequence into syllables using an automatic syllabification procedure." *Id.*, 1258(2:26-29), indicating that the syllable lexicon takes phones (i.e. phonemes) as input units. Because the syllable lexicon takes phonemes as an input to map to syllables, a skilled artisan would have understood the scored phonetic graph to output corresponding phonemes.

63. While Bazzi does not explicitly use the term "phoneme" one of ordinary skill would have understood that the phones of the received utterance derived by the

scored phonetic graph *P* satisfy the constitute "phonemes" as claimed by the '409 patent for a number of reasons. First, as explained with regard to the State of the Art in Section VI.B.1.a), above, one of ordinary skill understood that "phone" and "phoneme" were used interchangeably in the prior art literature and that, therefore, Bazzi's disclosure of phones corresponded to the claim's recitation of phonemes. Indeed, during prosecution of the '409 patent's counterpart EP application, the Applicant did not object to the Examiner's interpretation of Bazzi's disclosed use of "phone" as satisfying the claims' use of "phoneme." *See* Ex1006, 95 (arguing that, "[a]ssuming that the words phone and phoneme in this context are equivalent, [Bazzi] does not disclose using both phoneme and syllable recognition in the first … stage of interpretation").

64. Second, for the English language many phones and phonemes are the same. A phone constitutes the acoustic realization of a given phoneme, and a phoneme may—but many phonemes do not—have multiple allophones that may be expressed depending on context. As an example, English speakers generally express the phoneme /p/ differently in pronouncing the words "pin" and "spin," aspirating the "p" sound in the former but not the latter. EX1009, 48. In other words, each of the words "pin" and "spin" realizes a different allophone of the phoneme /p/. But many phones map to only a single phoneme in English. *See* EX1018, 436 (explaining that for "phonelike units," in "cases in which the acoustic and phonetic similarities

are roughly the same … then the phoneme and [phonelike unit] will be essentially identical."). Put another way, for any given phone, unless that phone can be mapped to multiple phonemes, recognition of that phone would also result in recognition of the corresponding phoneme. Accordingly, one of ordinary skill would understand that, by recognizing the probable phones of the input utterance via the scored phonetic graph, Bazzi would also recognize a string of phonemes corresponding to such phones.

65. Third, even if, for the sake of argument, one of ordinary skill would have not understood the "phones" disclosed by Bazzi to constitute "phonemes," it would have been an obvious design choice of the skilled artisan to implement the "phonetic units" of Bazzi's scored phonetic graph as phonemes instead of phones. It was known to one of ordinary skill that "recognition dictionaries often include a mix of phones and phonemes." EX1014, 2. And, as explained by Jurafsky, phonemes are often equated with the lexical level in the art, with lexicons thought of "as containing transcriptions expressed in terms of phonemes." EX1010, 104. Jurafsky presents two ways in which pronunciations of words can be transcribed at a lexical level: "When we are transcribing the pronunciations of words we can choose to represent them at this broad phonemic level; such a broad transcription leaves out a lot of predictable phonetic detail. We can also choose to use a narrow transcription that includes more detail, including allophonic variation." *Id*. It was therefore well

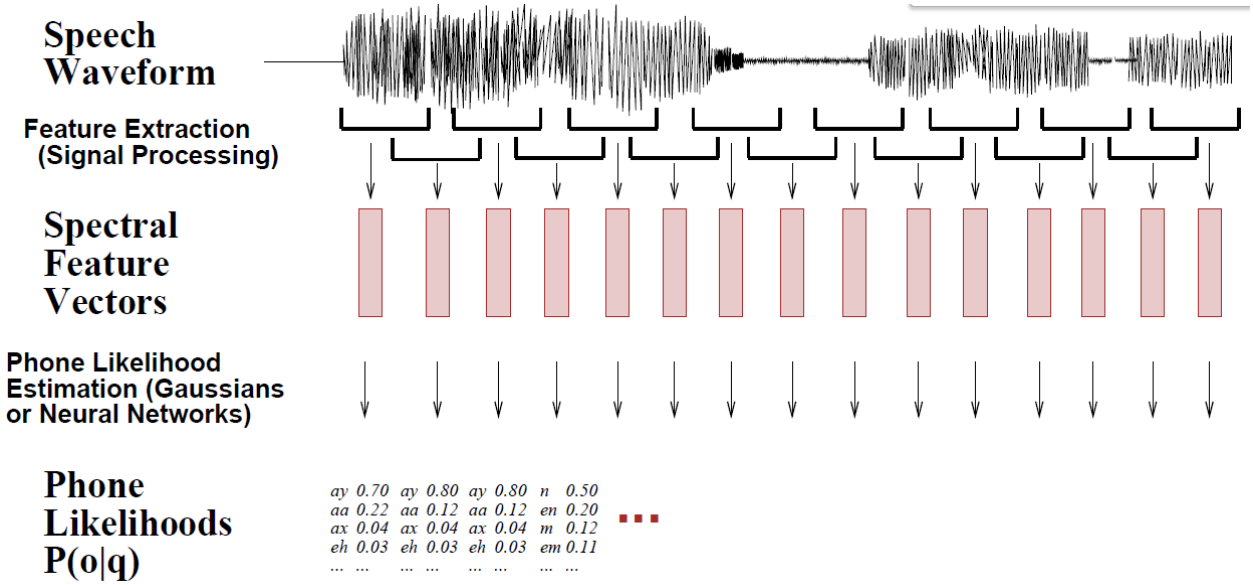
known to one of ordinary skill to present phonemes at the lexical level as an alternative to phones, and it would have been a simple design choice for the skilled artisan to generate the syllable lexicon ("L_s") based on phoneme-level transcriptions of the available lexicon, so as to map recognized phonemes to syllables. *See* EX1008, 1258(2:26-27) ("The syllable lexicon, L_s, is created from the word lexicon, L, through a direct mapping from phonetic units to syllabic units.").

66. Implementing the scored phonetic graph *P* as a graph of probable phonemes (as opposed to probable phones) to map to the phoneme-based L_s would have been well within the skill of one of ordinary skill in the art because acoustic models for modeling probable phonemes from an input acoustic signal were well known in the art. *See e.g.,* EX1018, 45-46 (describing performing feature measurement, detection, and segmentation to generate a "phoneme lattice" from which syllable and word lattices can be derived by integrating vocabulary and syntax constraints); EX1019, 2:42-3:2 (describing "prior art system" of generating acoustic model "which represent[s] each phoneme [] in [a] language" and using the acoustic model to link speech parameters in speech signal to phonemes); EX1020, Fig. 2, 5:14-40 (describing using a set acoustic models to "compute the probability of an acoustic sequence given a particular word sequence," with an acoustic model "used for each phoneme in [a] particular language."). Accordingly, implementing Bazzi such that its scored phonetic graph and syllable grammar utilize acoustic units of

phonemes instead of phones would have been an obvious design choice to one of ordinary skill. Therefore, one of ordinary skill would have understood that the scored phonetic graph of Bazzi represents a series of phonemes.

67. Moreover, Bazzi's generation and processing of the scored phonetic graph constitutes *recognizing* a phoneme stream. As explained above, the scored phonetic graph constitutes a graph structure representing the most likely phones contained in the received utterance. A skilled artisan would have understood such identification of the most likely phones to constitute "recognition" of those phones because in the field of automatic speech recognition (ASR), the term "recognizing" does not imply perfect certainty or final determination. Instead, it refers to the process by which the system analyzes an acoustic signal and produces a representation of the most likely linguistic units. This recognition process is inherently probabilistic, given the variable and noisy nature of spoken input. This understanding by one of ordinary skill is supported by Jurafsky, which, similar to Bazzi, describes a probabilistic graph structure of probable phones as the output of a "**phone recognition stage**." EX1010, 240-41 (emphasis added). Jurafsky explains that after initial signal processing of the input acoustic waveform, "we use statistical techniques like neural networks or Gaussian models to tentatively recognize individual speech sounds like p or b," and "the output of this stage is a vector of probabilities over phones for each frame." EX1010, 240. Jurafsky presents a

representation of such a graph of its phone likelihood estimation based on the input acoustic waveform in Figure 7.2, excerpted below:



EX1010, 241, Fig. 7.2. Accordingly, one of ordinary skill would understand Bazzi's generation of a scored phonetic graph to be consistent with the state of the art's "phone recognition stage" of identifying phone likelihoods in a received utterance. This understanding by one of ordinary skill is further confirmed by the '409 patent specification, which describes that recognizing a stream of phonemes may involve generating preliminary interpretations representing a set of best guesses. *See* EX1001, 5:65-6:6 ("[S]peech engine 112 may generate one or more preliminary interpretations of the user verbalization. The preliminary interpretations may represent a set of best guesses as to the user verbalization arranged in any predetermined form or data structure, such as an array, a matrix, or other forms. In one implementation of the invention, **speech engine 112 may generate the**

preliminary interpretations by performing phonetic dictation to recognize a stream of phonemes, instead of a stream of words.").

68. Finally, Bazzi discloses receiving the user's utterance and processing that utterance on an *electronic device*. Bazzi describes implementing its speech recognition in a "client/server architecture," where the "two-stage recognition process could be configured to have the first stage run locally on small client devices (e.g., hand-held portables) and thus potentially require less bandwidth to communicate with remote servers for the second stage." EX1008, 1258(1:8-10). Bazzi further discloses experimental testing of its two-stage recognizer system, which one of ordinary skill would have understood to be performed using a computing device programmed for that purpose. Indeed, a skilled artisan understood that, generally, speech recognitions systems were computerized systems that used electronics devices to carry out each of their various steps based on voice inputs from users. *See* EX1009, 5 (depicting and describing the "Basic system architecture of a speech recognition system."). For the reasons explained above, Bazzi's generation of a scored phonetic graph constitutes recognizing on an electronic device a stream of phonemes contained in the received utterance.

d) <u>Element [1.4]</u>

[1.4] mapping the recognized stream of phonemes to an acoustic grammar that phonemically represents one or more syllables, the recognized stream of

phonemes mapped to a series of one or more of the phonemically represented syllables; and

Bazzi teaches this limitation under the plain and ordinary meaning of "acoustic grammar"

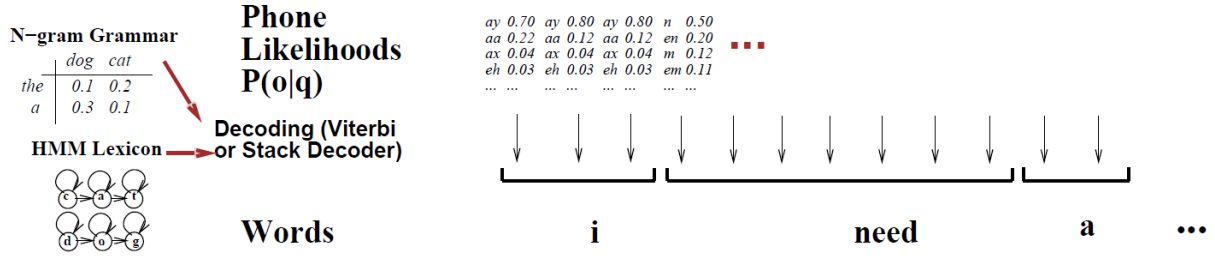
69. Bazzi discloses that the stream of phonemes of its scored phonetic graph, (denoted *P* in Bazzi's Equation 5), is mapped to an acoustic grammar that phonemically represents one or more syllables by composing the scored phonetic graph with a syllable lexicon and syllable grammar (denoted by $L_s \circ G_s$ in Bazzi's Equation 5). Bazzi's syllable recognizer, represented as an FST in Equation (5), includes a first stage as shown below:

$$S = (P \circ L_s \circ G_s) \circ (L_w \circ G) \qquad (5)$$

First Stage

EX1008, 1258(2:38) (Equation (5), annotated). Bazzi's Equation 5 reflects that in Bazzi's first stage the phonemes of the phonetic graph *P* are mapped to the syllable lexicon and grammar in order to map the recognized phonemes to corresponding syllables. The search space "*S*" is determined in two stages. "In the first stage we compute a syllable graph by searching the composition of *P* with the precomposed FST $L_s \circ G_s$." 1258(2:35-36). "L_s and G_s are the **syllable lexicon and grammar**,

respectively." *Id.*, 1258(2:26) (emphasis added). As explained for the State of the Art (section IV.B.3., above), lexicons and pronunciation grammars, such as those utilized by Bazzi, were well-known and widely used data structures in the art to transform a probabilistic representation of one phonetic unit to another. Jurafsky, for example, describes a conventional speech recognition system in which, in the system's decoding stage, "we take a dictionary of word pronunciations and a language model (probabilistic grammar)." EX1010, 241. This process is depicted in Jurafsky Figure 7.2, excerpted below, which shows that this dictionary of word pronunciations (a word-level lexicon in the form of an n-gram structure) and word-level grammar (in the form of a Hidden Markov Model structure) are applied to the probability graph of recognized likely phones to determine "the sequence of words which has the highest probability given the acoustic events." *Id.*, 241, Fig. 7.2.



70. Bazzi's syllable lexicon L_s is created from a word lexicon "through a **direct mapping from phonetic units to syllabic units**" by partitioning the phone sequence for each word in the lexicon "into syllables using an automatic syllabification procedure." EX1008, 1258(2:27-28). (emphasis added). As explained

in Section VIII.A.1.c), above, the relevant phonetic units of the syllable lexicon are phones, which one of ordinary skill would have understood to meet the '409 patent's recitation of "phonemes."[10] Accordingly, the syllable lexicon maps phonemes to corresponding syllables (representing those syllables phonemically), and the syllable lexicon thereby provides a model for mapping the recognized groups of phonemes in the scored phonetic graph to corresponding phonemically represented syllables.

71. Bazzi's grammar G_s constitutes "the syllable language model" that is built by starting with a "word-based training set" and "partition[ing] the words into syllables to obtain syllable sequences for training a syllable bigram or trigram." EX1008, 1258(2:27-28). Bigrams and trigrams are data structures that were well-known in the art and were commonly used to model probabilistic grammar. EX1015, 151; EX1010, 197-98. Specifically, bigrams and trigrams respectively model the probabilities of two or three units of speech (e.g., syllables in the case of syllable bigram/trigram) occurring in sequence. A skilled artisan would therefore have understood that the syllable grammar provides a language model reflecting the probability of a given syllable given the sequence of preceding syllables. One of

[10] Or alternatively, as explained in VIII.A.1.c), above, one of ordinary skill would have found it obvious to implement phonemes as the phonetic units used in Bazzi's system instead of phones.

ordinary skill would have understood that such "N-gram" language models were compatible with the sort of probabilistic finite state transducers (FSTs) used in Bazzi. Specifically, the methods all rely on using a prior sequence of units (e.g., phonemes, syllables, words) to estimate the probability of the next linguistic units in the sequence. When composed with the scored phonetic graph and lexicon, this probabilistic grammar (i.e., the n-gram model) helps determine the best candidates for mapping probable phonemes detected in the received utterance to the correct corresponding syllables by favoring or disfavoring syllables based on the observed probabilities of syllable sequences in the training data used to build the n-gram model.

72. The combination of the syllable lexicon and grammar thereby constitutes an "acoustic grammar that phonemically represents one or more syllables." Specifically, the syllable lexicon L_s provides the mapping from the scored phonetic graph P to syllables and the syllable grammar G_s provides the language model language model that helps to weigh potential interpretations based on recognized order of syllables. As explained above for element [1.3] (Section VIII.A.1.c), above) P represents Bazzi's recognized stream of phonemes. Bazzi maps this stream of phonemes to its acoustic grammar by "searching the composition of P with the precomposed FST $L_s \circ G_s$." EX1008, 1258(2:34-36). In searching the composition of the scored phonetic graph with the syllable lexicon and syllable

grammar, Bazzi maps the recognized phonemes to corresponding phonemically represented syllables contained in the acoustic grammar.

Bazzi teaches this limitation under the Amazon Litigation construction of "acoustic grammar"

73. Bazzi teaches this limitation even under the constructions of "acoustic grammar" applied in the Amazon Litigation and Amazon IPR. In the Amazon Litigation the parties agreed that the term "acoustic grammar" means "grammar of phonotactic rules of the English language that maps phonemes to syllables." EX1012, 2. Phonotactic rules, or phonotactic constraints, were well known in the art and refer to constraints in a language related to what phonetic units tend to co-occur or follow one another. *See* EX1010, 113 (describing "phonotactic constraint on what segments can follow each other"), EX1009, 730 ("A statistical model of phoneme co-occurrence, or phonotactics, was constructed over the training set."), EX1009, 151 ("The linguistic constraints employed by the second stage of this recognizer are based on the probabilities of groups of two or three words occurring in sequence."). As explained above in this Section, Bazzi's syllable grammar is based on the training of syllable bigrams or trigrams. Such syllable bigrams/trigrams constitute phonotactic rules because they provide an empirical model reflecting the phonotactics of the "word-based training set" (EX1008, 1258(2:31)) upon which the bigrams/trigrams are trained. When composed in Bazzi's FST at recognition, the

syllable grammar thereby constrains the determination of the probable paths through the composition based on the phonotactic rules reflected in the bigram/trigram model. Further, Bazzi explicitly contemplates its recognizer utilizing the English language. EX1008, 1257(1:33-38) (describing the problem in speech recognition caused by the constantly growing vocabulary of the English language). It would have been obvious for one of ordinary skill to implement Bazzi's system for English and implement its disclosed grammar for the English language including training its underlying models using English language dictionaries and training sets. Therefore, Bazzi's acoustic grammar (i.e., the combination of its syllable lexicon and syllable grammar) constitutes a grammar of phonotactic rules of the English language that maps phonemes to syllables.

Bazzi teaches this limitation under the Amazon IPR construction of "acoustic grammar"

74. In the Amazon IPR, the Board analyzed the specification of the '409 patent to construe the term "acoustic grammar" as it appeared in the claims of U.S. Patent No. 7,818,176 to mean "collection of the phonemes, or distinct units of sound of a spoken language, linked together to form syllables, which are linked together to form the words of the language." EX1013, 11-12. As explained above with regard to Bazzi's Equation (5), Bazzi's syllable lexicon is an FST corresponding to a graph that represents a series of phonemes linked together to form syllables. The syllable

grammar reflects the order of syllables in the language, i.e. the language model. With regard to linking syllables together to form the words of a language, Bazzi discloses a second recognition step, depicted in Bazzi's Equation (5), as annotated below.

$$S = (P \circ L_s \circ G_s) \circ (L_w \circ G) \qquad (5)$$

Second Stage

EX1008, 1258(2:38) (Equation (5), annotated). In this second stage, Bazzi applies a word lexicon and word grammar ("L_w" and "G") to the syllable graph output of the first stage. EX1008, 1258(2:36-37). The word lexicon and grammar perform analogous functions to the syllable lexicon and grammar of the first step, but instead of transforming the phonetic output of the scored phonetic graph *P* to corresponding syllables, the word lexicon and grammar transform the syllable units derived by the syllable lexicon and grammar to corresponding words. *Id.*, 1258(2:28-30) ("Entries in the second-stage word lexicon, L_w are represented by sequences of syllable units."). The word lexicon and word grammar therefore act to link the recognized syllables in the syllable graph to corresponding words to enable a best-path determination of the likely sequence of words in the utterance. *Id.*, 1257(1:17-19) ("A finite-state transducer speech recognizer is utilized to configure the recognition

as a two-stage process, where … syllable graphs are computed in the first stage, and passed to the second stage to determine the most likely word hypotheses."). Therefore, to the extent the Amazon IPR construction applies, the relevant "acoustic grammar" of Bazzi comprises the composition of Bazzi's syllable-level lexicon and grammar and word-level lexicon and grammar, because these structures include a collection of phonemes linked to form syllables (in the first stage of Bazzi's recognizer) and syllables linked together to form words (in the second stage of Bazzi's recognizer).

e) Element [1.5]

[1.5] generating at least one interpretation of the utterance, wherein the generated interpretation includes the series of syllables mapped to the recognized stream of phonemes.

75. Bazzi discloses this limitation because it generates one or more interpretations of the syllables contained in a received utterance in the form of a syllable graph. As explained in Sections VIII.A.1.c)-d), above, Bazzi determines a scored phonetic graph of the likely phonemes contained in a received utterance and, in the first stage of Bazzi's recognizer, transforms the phonetic units to corresponding syllables. The first-stage traversal of the syllable graph during recognition therefore represents tracing through the graph the likely syllables contained in the utterance, thereby providing one or more syllable-level interpretations of the utterance.

76. Bazzi also meets this limitation in a second way during the second stage in that it provides a word-level interpretation (or word-sequence level interpretation) of the received utterance. Again, Bazzi represents its two-stage, syllable-based recognizer using Equation (5):

$$S = (P \circ L_s \circ G_s) \circ (L_w \circ G) \qquad (5)$$

Second Stage

EX1008, 1258(2:38) (Equation (5), annotated). Bazzi's "second-stage search composes this FST with the precomposed word FST $L_w \circ G$," where L_w and G constitute the respective word lexicon and grammar, to produce a best word hypothesis. *Id.,* (2:36-37). Bazzi thereby searches the syllable graph generated in the first stage combined with the pre-generated word lexicon and grammar to determine word interpretations corresponding to the series of syllables contained in the syllable graph. *Id.,* 1257(1:17-19) ("A finite-state transducer speech recognizer is utilized to configure the recognition as a two-stage process, where … syllable graphs are computed in the first stage, and passed to the second stage **to determine the most likely word hypotheses**.") (emphasis added).

77. By outputting the word interpretation corresponding to the syllables contained in the syllable graph, which is generated by mapping Bazzi's phonetic

graph (i.e., the recognized stream of phonemes) to Bazzi's acoustic grammar (syllable lexicon and syllable grammar), Bazzi provides an interpretation that includes the series of syllables mapped to the recognized stream of phonemes.

2. Claim 2

> [2] The method of claim 1, the acoustic grammar phonemically representing the one or more syllables in accordance with acoustic elements of an acoustic speech model, wherein each syllable is represented by acoustic elements for an onset, a nucleus, and a coda.

78. As explained in Section VIII.A.1.d), above , the acoustic grammar of Bazzi includes the syllable lexicon, L_s. And as explained in that section, the syllable lexicon includes acoustic elements of both phonemes and syllables and phonemically represents one or more syllables by providing a mapping of syllables to corresponding phonemes contained in the syllables. This phonemic representation of syllables is in accordance with an acoustic speech model of a language as defined by the training set used for training Bazzi's acoustic grammar. *See* Bazzi, 1258(2:26-27) (explaining that the syllable lexicon is created from a word lexicon). For example, as explained in Section VIII.A.1.d) , above, one of ordinary skill would have found it obvious to implement the system of Bazzi for the English language and would have therefore selected an appropriate word lexicon and word-based training set to provide the acoustic speech model for the English language upon which to build the acoustic grammar. As such, Bazzi's acoustic grammar (the

syllable lexicon and syllable grammar) includes acoustic elements (i.e. phonemes and syllables) as used in a language and that phonemically represents one or more syllables in accordance with the acoustic elements of an acoustic speech model of that language.

79. With regard to the claim's recitation of "wherein each syllable is represented by acoustic elements for an onset, a nucleus, and a coda," one of ordinary skill would have understood Bazzi's syllable lexicon would have included such a representation. A skilled artisan would have been aware that the onset, nucleus, and coda are nothing more than fundamental phonetic components of a syllable. *See, e.g.,* EX1001, 2:50-52 ("Portions of a word may be represented by a syllable, which may be further broken down into core components of an onset, a nucleus, and a coda."); EX1010, 102 ("A syllable is usually described as having an optional initial consonant or set of consonants called the onset, followed by a vowel or vowels, followed by a final consonant or sequence of consonants called the coda."). Because Bazzi's syllable lexicon represents syllables phonetically, i.e., it represents what constituent phonemes make up a given syllable, one of ordinary skill would recognize that those constituent phonemes also reflect the core components—the onset, nucleus, and coda—of the syllable. That is, one of ordinary skill would recognize the initial consonant phonemes in the syllable lexicon for a given syllable reflect the syllable's onset, the vowel phoneme in the syllable lexicon for the syllable

reflects the nucleus, and the consonant phonemes following the vowel in the lexicon for the given syllable reflect the coda. In summary, because Bazzi's syllable lexicon maps syllables to their constituent phonemes, one of ordinary skill would have understood that those phonemes comprise the onset, nucleus, and coda components of the syllable.

3. Claim 3

[3] The method of claim 2, the acoustic grammar including transitions between the acoustic elements, wherein the transitions are constrained according to phonotactic rules of the acoustic speech model.

80. Bazzi discloses this limitation. As explained in Section VIII.A.1.d), above, Bazzi's acoustic grammar (syllable lexicon L_s and syllable grammar G_s) are trained using a word lexicon and word-based training set that model the rules the language. The acoustic grammar therefore reflects an acoustic speech model for a given language upon which the acoustic grammar is trained. The acoustic grammar of Bazzi includes transitions between its acoustic elements; specifically transitions between recognized phonemes and corresponding syllables. *See* Section VIII.A.1.d), above (explaining how Bazzi's acoustic grammar maps phonemes to syllables). This is evidenced by the fact that Bazzi's first stage takes as input the scored phonetic graph representing recognized likely phonemes and composes that graph to produce a syllable graph. EX1008, 1258 (2:22-35). Bazzi's acoustic grammar further

contains phonotactic rules[11] that constrain such transitions. Bazzi's syllable lexicon, which maps phonemes to syllables, constrains such phoneme-to-syllable transition because the syllable lexicon dictates which syllables correspond to the stream of phonemes recognized in the received utterance. EX1008, 1258(2:22-28). The syllable lexicon thereby constrains allowable combinations of phonemes in the syllables of a given language.

81. Bazzi's acoustic grammar also includes inter-acoustic element transitions. That is, Bazzi's syllable grammar includes phonotactic rules in syllable-to-syllable transitions: As explained in Section VIII.A.1.d),above, Bazzi's syllable grammar (i.e. "the syllable language model" (EX1008(2:31-34)) constitutes a trained model of syllable bigrams or trigrams to constrain the allowable sequence of syllables; syllable bigrams reflect the probability of two given syllables being in sequence, while syllable trigrams reflect the sequential probability for three given syllables. These syllable bigrams/trigrams provide a model that specifies the relative probability of a given syllable in a language given the preceding syllable. Accordingly, the syllable bigrams/trigrams of the syllable grammar constitute a

[11] As explained in Section VIII.A.1.d), , above, phonotactic rules refer to constraints in a language governing the allowable sequences of phonetic elements (e.g., allowable syllable structures or phoneme combinations).

phonotactic rule constraining the transitions of one syllable to another by restricting allowable syllable sequences.

4. Claim 6

a) Element [6.1]

[6.1] The method of claim 1, further comprising: generating a plurality of candidate interpretations of the utterance, wherein each candidate interpretation includes a series of words or phrases corresponding to the series of syllables mapped to the recognized stream of phonemes;

82. Bazzi discloses this limitation. As explained in Sections VIII.A.1.c)-d),, above, the first stage of Bazzi's recognizer composes the scored phonetic graph (i.e., a graph structure of likely phonemes recognized in the received user utterance) with a syllable lexicon and syllable grammar to map the probable phoneme series of the scored phonetic graph to syllables, thereby generating a syllable graph. Bazzi's second recognition stage further composes this output with a word-level lexicon and grammar to apply further constraints as well as recognize the words corresponding to the series of syllables contained in the syllable graph. Thus, in the Bazzi recognizer, "syllable graphs are computed in the first stage, and passed to the second stage **to determine the most likely word hypotheses**." EX1008, 1257(1:18-19) (emphasis added).

83. Bazzi describes that in its FST-based framework, "recognition can be viewed as **finding the best path(s) in the composition**." *Id.*, 1258(1:28-29)

(emphasis added). . Bazzi explicitly describes finding multiple potential best paths through its two-stage recognition process using the FST expressed in equation (5). The FST composed of the phonetic graph, syllable lexicon and grammar, and word lexicon and grammar applies constraints at each level to the recognition process (*see* Section VI.B.4., above, describing FST operation). This results in multiple candidate paths that are more or less probable given the weights associated with that path. Each complete path from start to end in the combined graph represents a sequence of phonemes, syllables, and words; i.e. a candidate interpretation. As explained in Section VII.A.1.d), above, the syllable lexicon provides the mappings from each syllable to a set of phonemes, and therefore the mappings from a series of syllables to a series of phonemes and during recognition from the series of *recognized* phonemes to a series of *recognized* syllables. When the second-stage lexicon and grammar are included, every candidate interpretation thus includes "a series of words or phrases corresponding to the series of syllables mapped to the recognized stream of phonemes" and therefore Bazzi discloses this claim element.

84. Further, various algorithms were known in the for finding such best paths and Bazzi identifies two such algorithms: "Typical recognizer configurations deploy a bigram language model in a forward Viterbi search, while a trigram (or higher-order) language model is used in a backward A* search." *Id*. These same algorithms were well known to a skilled artisan. Jurafsky, for example describes

their use in conventional speech recognition systems: "Finally, in the decoding stage, we take a dictionary of word pronunciations and a language model (probabilistic grammar) and use a Viterbi or A* decoder **to find the sequence of words which has the highest probability given the acoustic events**." EX1010, 241 (emphasis added); *see also* EX1009, 592 ("Speech recognition search is usually done with the Viterbi or A* stack decoders."). And using a forward Viterbi search algorithm together with a backward A* search to find best paths for word sequences, as Bazzi suggests was also well known in the art. *See* EX1009, 670-71 (describing the "Forward-Backward Search Algorithm" using a Viterbi search for forward searching and A* search for backward searching).

85. Accordingly, it would have been obvious to one of ordinary skill to utilize the search algorithms (i.e., a Viterbi forward search and A* backward search) to implement Bazzi's search methodology. As explained further below, a person of ordinary skill in the art recognized that such a search methodology generates and scores a plurality of candidates for a given utterance.

b) <u>Element [6.2]</u>

[6.2] assigning a score to each of the plurality of candidate interpretations; and

86. As explained for Element [6.1] Bazzi teaches that "recognition can be viewed as **finding the best path(s) in the composition**." And as explained for Element [6.1], a person of ordinary skill in the art would implement the speech

recognition using a forward-backward search methodology using a forward Viterbi search and backward A* search and Bazzi discloses this limitation through its use of such a methodology.

87. Using such a forward-backward search methodology, "[t]he idea is to first perform a forward search, during which partial forward scores α for each state can be stored." EX1009, 670. Subsequently, a backward A* search (also referred to in the art as stack decoding is performed) whereby "the first complete hypothesis found with a cost below that of all the hypotheses in the stack is guaranteed to be the best word sequence." *Id*., 671. Thereafter, subsequent complete hypotheses correspond sequentially to the *n*-best list, as they are **generated in increasing order of cost**." EX1009, 672 (emphasis added). Therefore, Bazzi's disclosed forward-backward search methodology results in scoring the plurality of candidate interpretations. Use of such a backward A* search achieves Bazzi's purpose of finding multiple best paths. *Id*., 1008, 1258(1:28-29) ("[R]ecognition can be viewed as finding the best path(s) in the composition."). As Huang explains, "It is straightforward to extend stack decoding to produce the *n*-best hypotheses by continuing to extend the partial hypotheses according to the same A* criterion until *n* different hypotheses are found. These *n* different hypotheses are destined to be the *n*-best hypotheses." EX1009, 671. Accordingly, a person of ordinary skill would have implemented Bazzi's disclosed forward-backward search methodology in

Bazzi's speech recognition system to find the scored best paths for a given composition.

c) Element [6.3]

[6.3] selecting a candidate interpretation having a highest assigned score as being a probable interpretation of the utterance.

88. As explained above for Element [6.2], one of ordinary skill would have found it obvious to use Bazzi's forward-backward search method to find the best paths through the composition of Bazzi's recognition model, thereby resulting in an n-best list of hypothesized word sequence candidates for a received utterance. Further, a skilled artisan would have understood that this methodology also results in selection of a candidate having a highest score as the best interpretation. Regarding use of the A* backward search algorithm, Huang explains that "**[t]he first complete hypothesis generated by backward A* search** coincides with the best one found in the time-synchronous forward search and **is truly the best hypothesis**. Subsequent complete hypotheses correspond sequentially to the *n*-best list, as they are **generated in increasing order of cost**." EX1009, 672 (emphasis added). Thereby, Bazzi's use of the backward A* algorithm results in an *n*-best list of word sequence hypotheses sorted by cost with the lowest cost hypotheses corresponding to the best probable interpretation of the utterance's word sequence. The skilled artisan would have understood a lowest-cost word sequence interpretation (as

returned by the backward A* algorithm) and a "highest assigned scored" to be equivalent because they would have recognized the path having the lowest cost to have the highest probability of being the correct hypothesis. It would have been evident to that POSITA that a cost is mathematically interchangeable with an inversely proportional score because the skilled artisan understood that a lower cost for a word sequence reflects a higher probability of that word sequence being a correct hypothesis. *See e.g.,* EX1010, 187; EX1021 7:57-67.

89. Therefore, It would have been an obvious design choice for one of ordinary skill to implement Bazzi to identify the best word hypothesis based on a highest probability for path (i.e., a score) instead of as a lowest cost, because a POSITA would have recognized a score or a cost as mathematically interchangeable and constituting a limited number of choices for mathematically presenting the quality of a searched word hypothesis. In the field of speech recognition it was well known to represent a best interpretation hypothesis as either a highest probability (i.e., a probability score) or as a weight calculated as the negative log probability. *See e.g.,* EX1010, 187 ("As is commonly true with probabilistic algorithms, they actually use the negative log probability of the word (-log ($P(w)$)."); EX1021, 7:57-67 ("'Score' is a numerical evaluation of how well a given hypothesis matches some set of observations. Depending on the conventions in a particular implementation, better matches might be represented by higher scores (such as with probabilities or

logarithms of probabilities) or by lower scores (such as with negative log probabilities or spectral distances"); EX1022, [0030] (same); EX1023, 3:52-60 (chart showing probability of phone realizations as both "probability" and "weight = -log prob" and showing how the highest scored probability correlates to the lowest weight). Further, one of ordinary skill in the art would have understood that the FSTs taught by Bazzi are probabilistic and that therefore selecting the best path would have been equivalent to that with the highest probability, i.e., the highest score. The '409 patent itself recognizes that a best word hypothesis can be interchangeably determined using a highest or lowest score. EX1001, 10:52-54 ("In one implementation of the invention, a candidate interpretation with a highest (or lowest) score may be designated as a probable interpretation."). Thus, it would be an obvious design choice to one of ordinary skill implementing the system of Bazzi to represent word hypothesis paths using a score that reflected the probability that a path corresponds to the best interpretation, rather than implementing the best path as the lowest cost path.

IX. GROUND 2: CLAIMS 2 & 3 ARE OBVIOUS OVER BAZZI IN FURTHER VIEW OF SABOURIN

90. As explained below, it is my opinion that claims 2 and 3 are obvious over Bazzi in further view of Sabourin.

A. Sabourin

91. Sabourin is U.S. Patent No. 6,108,627 titled "AUTOMATIC TRANSCRIPTION TOOL." EX1024, Title Page. Sabourin was filed on October 31, 1997, and issued on August 22, 2000. *Id*. I understand that Sabourin is therefore prior art under at least 35 U.S.C. § 102(b).

92. Sabourin describes a method for phonemic transcription and generation of phonemic transcription dictionaries for use in speech recognition systems. Sabourin states the following:

> A "phonemic transcription" encodes the sound patterns of a word using the phonemic alphabet. In addition to symbols from the phonemic alphabet, phonemic transcriptions may additionally include information relating to word stress and syllabification…. Phonemic transcription dictionaries are useful in a number of areas of speech processing, such as in speech recognition.

Id., 1:31-35, 41-43.

93. Sabourin discloses "[a]n automatic transcription tool … us[ing] a variety of transcription methods to generate relatively accurate phonemic transcriptions." *Id.*, 1:60-62. In addition to this phonemic transcription, Sabourin also discloses "automatically partition[ing] a transcription into syllables." *Id*., 2:35-

36. Sabourin's method for performing syllabification is shown in its Figure 10, reproduced below. *Id*. 13:2-4.

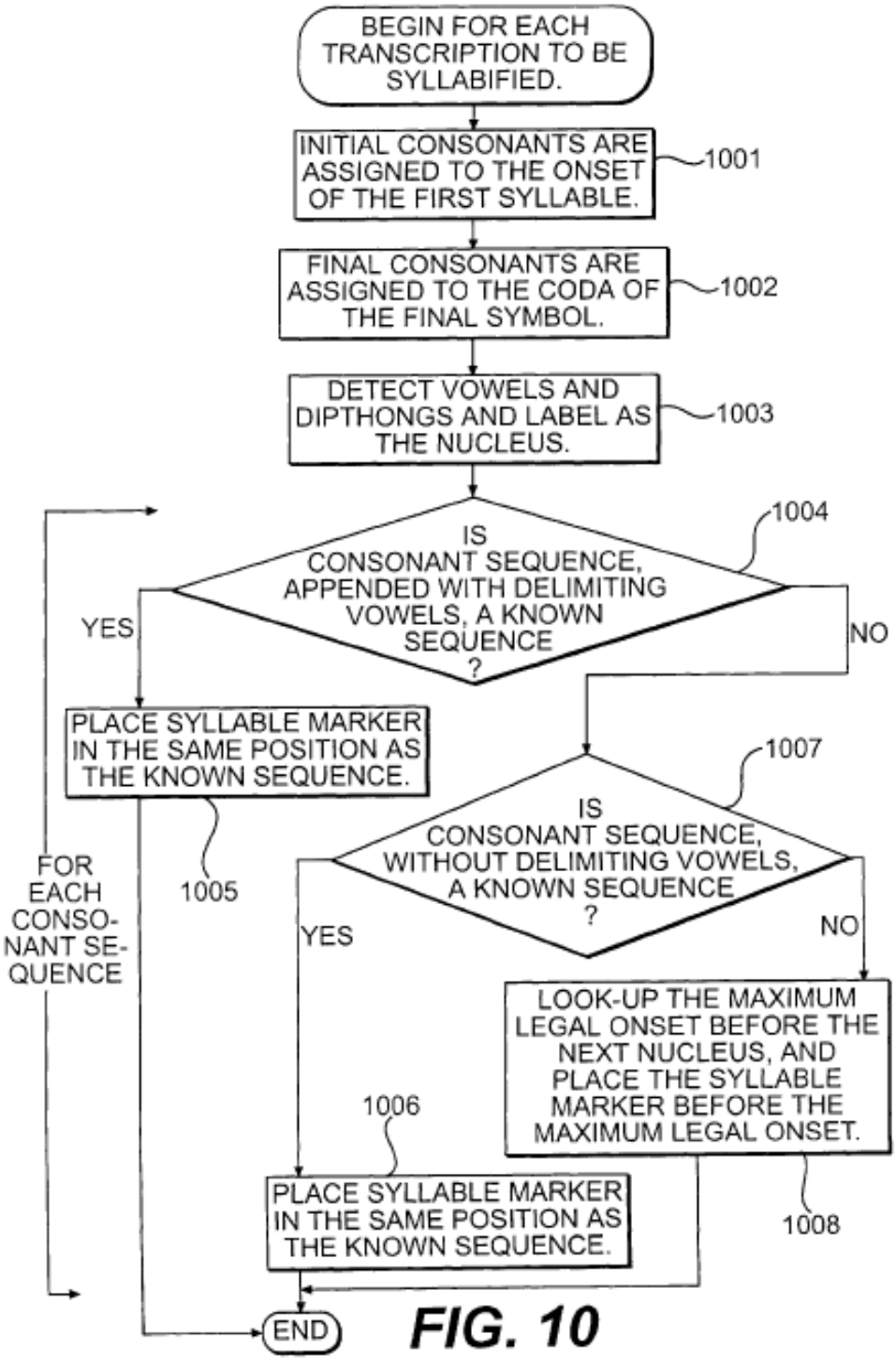


FIG. 10

Id., Fig. 10.

94. Sabourin performs automatic phoneme transcription by first generating a grapheme (i.e. letter) mapping from a training dictionary and then assigns a mapping value to each mapped grapheme-to-phoneme pair based on the relative frequency with which a particular phoneme string corresponds to its associated grapheme string. *Id*., Fig. 5, 9:19-10:22. Then a phonemic transcription is created for each word in the training dictionary by decomposing each input orthography (i.e., word) into possible component substrings, using the assigned grapheme-to-phoneme mapping values to generate a transcription score, and selecting the component substring decomposition with the highest score as the best transcription hypothesis. *Id*., Fig. 6, 10:24-38.

95. Following its phonemic transcription process, Sabourin performs word transcription post-processing including syllabification, stress assignment, and phonotactic post-processing. *Id*., 10:54-60. The first transcription post-processing step is "automatically partition[ing] a transcription into syllables." *Id*., 2:35-36. Regarding this syllabification procedure, Sabourin describes the following:

> For each input transcription to be syllabified, syllabification section 802 begins by assigning initial consonants to the onset of the first syllable (step 1001). Similarly, final consonants are assigned to the coda of the

> final symbol (step 1002). Vowels and diphthongs are then detected and labeled as nuclei (step 1003).

96. *Id*., 13:5-10. In addition to adding syllabification information for transcribed phonemes, Sabourin discloses assigning stress information to syllables. *Id*., 13:46-53. Lexical stress refers to the amount of energy expressed in a syllable, with stressed syllables being pronounced louder or longer. EX1010, 103. Finally, following syllable stress assignment, Sabourin discloses performing "phonotactic post-processing" on the syllabified, stress assigned transcriptions generated according to Sabourin's method to verify and prune the generated transcriptions. EX1024, 15:8-9. Sabourin explains:

> Phonotactic validation is the process of verifying the generated transcriptions. Preferably, for English transcriptions, the following phonotactics are checked: lax-tense vowel combinations, invalid consonant sequences, implausible vowel beginning or endings, implausible consonant beginnings or endings, double phonemes, and single syllable transcriptions whose only vowel is a schwa. These phonotactic checking algorithms, as well as other possible ones, are rule-based, and are all within the capabilities of one of ordinary skill in the art. If phonotactic irregularities are detected, the transcription is labeled as being phonotactically illegal and is aborted.

Id. 15:9-19.

B. <u>The Bazzi-Sabourin Combination</u>

97. One of ordinary skill would have been motivated to combine the teachings of Bazzi and Sabourin. Specifically, a skilled artisan would have been motivated to utilize Sabourin's automatic phonemic transcription method, including automatic syllabification, to generate the syllable lexicon of Bazzi.

98. Bazzi discloses that for "each word in the [word] lexicon, we partition the phone sequence into syllables using an automatic syllabification procedure." EX1008, 1258(2:27-28). Bazzi does not provide details of how "syllabification" is accomplished, and a skilled artisan would recognize that this process would be critical to creating the syllable lexicon. Accordingly, one of ordinary skill would look toward a suitable automatic syllabification procedure in order to partition phone sequences into syllables. Sabourin discloses just such a technique in that it discloses "automatically partition[ing] a [phonemic] transcription into syllables." EX1024, 2:36. One of ordinary skill would recognize the advantages of using Sabourin's automatic syllabification technique, including labeling the phonemes corresponding to the onset, nucleus, and coda. One of ordinary skill would also have recognized that Sabourin's disclosure of assigning stress to transcribed syllables could improve word recognition performance, because "difference in lexical stress can affect the meaning of a word," such as the noun "content" and the adjective "content."

EX1010, 103. Therefore, Sabourin's identification of lexical stress in the generated lexicon would help in correctly distinguishing words that may differ based on such stress. Finally, one of ordinary skill would also have implemented Sabourin's phonemic transcription methodology for the benefit of its disclosed phonotactic post-processing in order to provide improved verification of the legality of transcriptions in the lexicon. The skilled artisan would understand that by implementing such phonotactic post-processing, the lexicon would include fewer invalid transcriptions, thereby potentially making the speech recognition system both more efficient and more accurate.

99. One of ordinary skill would have had a reasonable expectation of success in combining Bazzi and Sabourin because Bazzi explicitly calls for an automatic syllabification procedure to generate a "mapping of phonetic units to syllabic units" (EX1024, 1258) and Sabourin teaches just such a procedure. Moreover, pronunciation dictionaries (i.e., lexicons) such as those generated by Sabourin that "include[ed] syllabification and stress" were well known in the art (EX1010, 135), and one of ordinary skill would therefore understand such a lexicon to be usable as the syllable lexicon as disclosed by Bazzi.

1. Claim 1

100. The combination of Bazzi and Sabourin renders claim 1 of the ’409 patent obvious for the same reasons as Bazzi alone, as explained in Section VIII.A.1., above.

2. Claim 2

> [2] The method of claim 1, the acoustic grammar phonemically representing the one or more syllables in accordance with acoustic elements of an acoustic speech model, wherein each syllable is represented by acoustic elements for an onset, a nucleus, and a coda.

101. As explained in VIII.A.1.d), above, Bazzi teaches an acoustic grammar, including its syllable lexicon and syllable grammar. And as explained above in Section IX.B., above, it would have been obvious to one of ordinary skill to use the phonemic transcription and syllabification methodology of Sabourin to generate the syllable lexicon of Bazzi.

102. Sabourin teaches that syllables “are defined” as a structure made up of onset, nucleus, and coda:

> Syllables relate to the rhythm of a language, and, as used in this disclosure, are defined as a collection of phonemes with the following structure:
>
> [onset] nucleus [coda]
>
> where the brackets around “onset' and “coda” indicate that these components are optional. The “onset' and

> "coda" are a sequence of one or more consonants, and the nucleus is a vowel or diphthong.

103. Further, as Sabourin discloses, its phonemic transcription includes syllabification and explicitly labels the onset, nucleus, and coda components of mapped syllables:

> For each input transcription to be syllabified, syllabification section 802 begins by **assigning initial consonants to the onset** of the first syllable (step 1001). Similarly, **final consonants are assigned to the coda** of the final symbol (step 1002). **Vowels and diphthongs are then detected and labeled as nuclei** (step 1003).

104. *Id*., 13:5-10 (emphasis added). Accordingly, the syllable lexicon that would have been generated using Sabourin's phonemic transcription and syllabification includes each syllable represented by acoustic elements for an onset, a nucleus, and a coda. Those syllables are in accordance with acoustic elements (phonemes and syllables) of an acoustic speech model provided by Sabourin. That is, Sabourin's acoustic speech model is its methodology for phoneme transcription and post-transcription processing using a training dictionary for a language to be modeled. Therefore, the combination of Sabourin's speech model, including the syllabification process, with Bazzi's syllable lexicon, would have rendered Claim 2 obvious.

3. Claim 3

[3] The method of claim 2, the acoustic grammar including transitions between the acoustic elements, wherein the transitions are constrained according to phonotactic rules of the acoustic speech model.

105. As explained above in IX.B., above, it would have been obvious to one of ordinary skill to use the phonemic transcription and syllabification methodology of Sabourin to generate the syllable lexicon of Bazzi. This methodology includes Sabourin's "phonotactic post-processing," which constrains transitions between acoustic elements, by checking for and discarding transcriptions having the following phonotactic irregularities: "lax-tense vowel combinations, invalid consonant sequences, implausible vowel beginning or endings, implausible consonant beginnings or endings, double phonemes, and single syllable transcriptions whose only vowel is a schwa." EX1024, 15:9-19. Accordingly, the transitions between acoustic elements (i.e. phonemes and syllables) in the syllable lexicon (and therefore the acoustic grammar comprising the syllable lexicon) generated using Sabourin's phonemic transcription and syllabification, is constrained according to the phonotactic rules of the acoustic speech model provided by Sabourin (i.e. Sabourin's methodology for phoneme transcription and post-transcription processing).

106. While Bazzi does not use the word "phonotactic", Sabourin does, and adds phonotactic post-processing as a separate stage of transcription. This meets

Claim 3's limitation of "acoustic grammar including transitions … constrained according to phonotactic rules" of an acoustic speech model. Therefore, it is my opinion that Claim 3 would have been obvious to one of ordinary skill over Bazzi in light of Sabourin.

X. GROUND 3: CLAIM 6 IS OBVIOUS OVER BAZZI IN FURTHER VIEW OF EPSTEIN

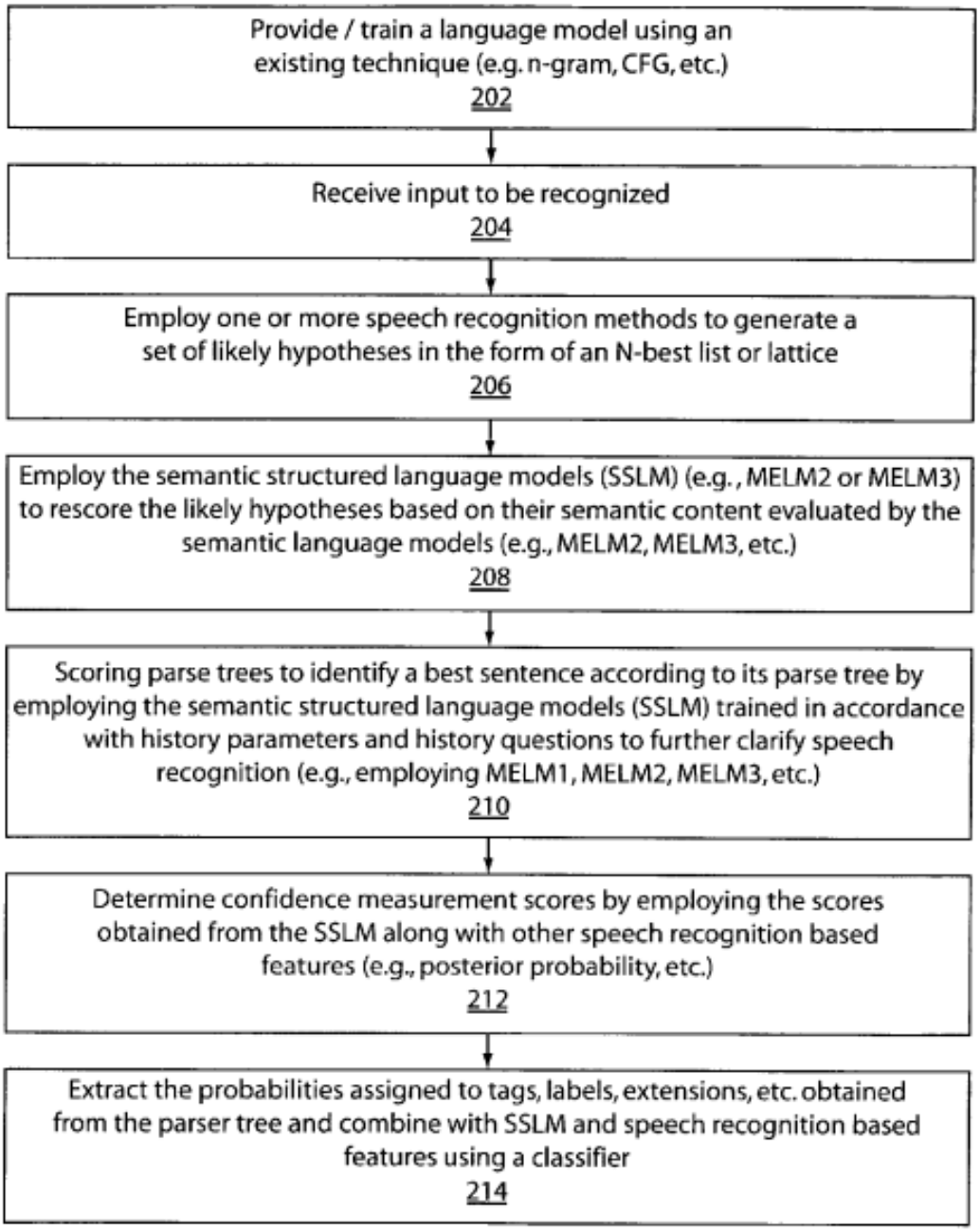
A. Epstein

107. Epstein is U.S. Patent Publication No. 2005/0055209A1 titled "Semantic language modeling and confidence measurement." EX1025, Title Page. Epstein was filed on May 9, 2003, and published on March 10, 2005. *Id*. I understand that Sabourin is therefore prior art under at least 35 U.S.C. §§ 102(a) and (b).

108. Epstein describes a "system and method for speech recognition [that] includes generating a set of likely hypotheses in recognizing speech, rescoring the likely hypotheses by using semantic content by employing semantic structured language models, and scoring parse trees to identify a best sentence according to the sentence's parse tree by employing the semantic structured language models to clarify the recognized speech." *Id.*, Abstract. Epstein describes existing problems with speech recognition systems relying only on conventional n-gram language models: "Although n-gram language models achieve a certain level of performance, they are not optimal. N-grams do not model the long-range dependencies, semantic

and syntactic structure of a sentence accurately." *Id*., [0005]. Epstein addresses this problem by employing a second stage after initial candidate recognition to re-score candidate sentence interpretations based on a model taking into account sentence semantics (a "semantic structured language model"). *Id*., [0018]. Epstein's semantic structure language model reflects a model of semantic information for a language, which "may include one or more of word choice, order of words, proximity to other related words, idiomatic expressions or any other information based word, tag, label, extension or token history." *Id*. [0028].

109. Epstein's speech recognition method is depicted in its Figure 4, reproduced below:



EX1025, Fig. 4. As Epstein explains, following receipt of a user speech input (at block 204), the method then (at block 206) employs "one or more speech recognition methods … to generate a set of likely hypotheses. The hypotheses are preferably in

the form of an N-best list or lattice structure." EX1025, [0068]. These likely hypotheses are then rescored in block 208 using semantic structured language models (SSLM) to "rescore the likely hypotheses based on the semantic content of the hypotheses." *Id*., [0069]. Finally, the best sentence interpretation is identified by scoring parse trees of the sentence hypotheses using the semantic structured language models. *Id*. A parse tree is a hierarchical representation of the structure of a sentence according to the rules of a grammar. *See* EX1009, 62; EX1025, [0023], [0045], Fig. 2.

A. The Bazzi-Epstein Combination

110. One of ordinary skill in the art would have been motivated to combine the teachings of Bazzi and Epstein. Specifically, one of ordinary skill would have been motivated to implement the speech recognition methods of Bazzi to generate an initial set of likely sentence hypotheses and then to apply Epstein's teachings to re-score those hypotheses and select the best sentence interpretation using Epstein's semantic language models.

111. One of ordinary skill would have been motivated to make such combination due to the known limitations of N-gram language models, such as those implemented by Bazzi (*see* EX1008, 1258(2:31-34) (describing use of syllable bigrams and trigrams)), and the well-known addition of semantic models such as taught by Epstein to overcome those limitations. Epstein itself provides this

motivation by explaining that N-gram language models are not optimal, in that they "do not model the long-range dependencies, semantic and syntactic structure of a sentence accurately." EX1025, [0005]. And Epstein explains that its semantic language modeling techniques "improve speech recognition accuracy." Indeed, implementing a system that iteratively determines an initial set of best hypotheses using a less sophisticated, more efficient knowledge source (e.g., a bigram language model such as used in Bazzi) and then rescores those hypotheses using a more sophisticated knowledge source (e.g., a semantic language model such as in Epstein) was well known in the art. Such a system is shown, for example in Figure 7.2 of Jurafsky, reproduced below.

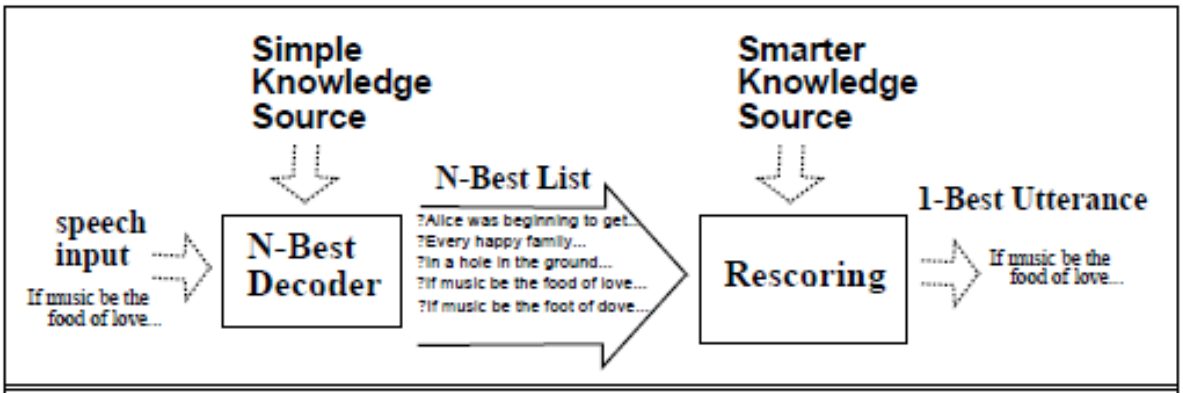


Figure 7.12 The use of *N*-best decoding as part of a two-stage decoding model. Efficient but unsophisticated knowledge sources are used to return the *N*-best utterances. This significantly reduces the search space for the second pass models, which are thus free to be very sophisticated but slow.

EX1010, 253. Schwartz, et al., describe a similar implementation of applying a more efficient set of knowledge sources, including statistical grammar to generate a list of

sentence candidates, and then re-ordering that list using additional knowledges sources, including semantics. EX1026, 81-82, Fig. 1; *see also* EX1027, 2:26-47 (describing a speech recognition paradigm of using lesser cost knowledge sources to output "a list of the most likely whole sentence hypotheses" and rescoring list using remaining knowledge sources to find the "highest overall scoring sentence" and stating that "[t]his approach has produced some impressive results"). Accordingly, one of ordinary skill would have been motivated to combine the teachings of Bazzi and Epstein because one of ordinary skill would have understood such combined speech recognition system to provide potentially improved speech recognition accuracy (by rescoring initial hypotheses using more sophisticated knowledge sources) while not requiring significantly more computational resources (by reducing the search space to the hypotheses returned by the initial search).

112. One of ordinary skill would have had a reasonable expectation of success in combining the teachings of Bazzi and Epstein. As explained in the preceding paragraph, two-stage searches performing rescoring of an initial search using a smarter knowledge source (including semantic knowledge) were well known in the art, and one of ordinary skill would have therefore expected to be successful in using Bazzi's speech recognition method as such an initial search in combination with Epstein's teachings regarding reordering using a semantic language model. Moreover, Epstein does not specify a particular speech recognition method for

generating initial hypotheses, only stating that "one or more speech recognition methods may be employed to generate a set of likely hypotheses." EX1025, [0068]. Bazzi performs just such a speech recognition method of generating a set of likely hypotheses. EX1008, 1257(1:19) (determining "the most likely word hypotheses"), EX1258(1:28-29) ("[R]ecognition can be viewed as finding the best path(s).").

1. Claim 1

113. The combination of Bazzi and Epstein renders claim 1 of the '409 patent obvious for the same reasons as Bazzi alone, as explained in Section VII.A., above.

2. Claim 6

a) Element [6.1]

[6.1] The method of claim 1, further comprising: generating a plurality of candidate interpretations of the utterance, wherein each candidate interpretation includes a series of words or phrases corresponding to the series of syllables mapped to the recognized stream of phonemes;

114. The Bazzi-Epstein combination teaches this limitation for the same reasons as Bazzi alone, as explained in Section VIII.A.4., above. That is, Bazzi explicitly discloses generating a plurality of best path hypotheses for an utterance consisting of a series of words, where those words correspond to the syllables mapped to the recognized stream of phonemes through Bazzi's speech recognition method. Additionally, and in the alternative, as explained in that Section, one of

ordinary skill would have found it obvious to implement the search algorithms disclosed by Bazzi to output its word series hypotheses in the form of an n-best list. One of ordinary skill would be further motivated to do so in combining Bazzi with Epstein to match Epstein's preferred format. EX1025, 7:38-42 ("[O]ne or more speech recognition methods may be employed to generate a set of likely hypotheses. **The hypotheses are preferably in the form of an N-best list** or lattice structure.") (emphasis added).

b) Element [6.2]

[6.2] assigning a score to each of the plurality of candidate interpretations; and

115. The Bazzi-Epstein combination teaches this limitation. After performing an initial speech recognition to determine a set of likely sentence hypotheses (performed by the speech recognition method of Bazzi in the Bazzi-Epstein combination), Epstein explicitly discloses assigning scores to each such likely sentence hypothesis by **rescoring** them using Epstein's semantic language models. EX1025, 7:43-48. ("In block 208, semantic structured language models (SSLM) are employed to **rescore** the likely hypotheses based on the semantic content of the hypotheses. In block 210, parse trees are scored to identify a best sentence in accordance with its parse tree. This is performed by using SSLMs.")

(emphasis added). The Bazzi-Epstein combination therefore assigns a score to each of a plurality of candidate sentence interpretations.

c) Element [6.3]

[6.3] selecting a candidate interpretation having a highest assigned score as being a probable interpretation of the utterance.

116. The Bazzi-Epstein combination teaches this limitation. Epstein explicitly states that a best sentence interpretation is selected based on the score for that sentence determined during its rescoring procedure. EX1025, 7:47-48 ("In block 210, parse trees are scored to identify a best sentence in accordance with its parse tree."). While Epstein does not explicitly state that the score for the best sentence is the "highest assigned score," it would have been obvious for one of ordinary skill to implement Epstein such that its best sentence hypothesis is determined using a highest assigned score. First, it would be common sense for one of ordinary skill to score an ordered list of best interpretations from best to worst using a highest-to-lowest score. Further, it would have been an obvious design choice because it was well known in the art that best interpretations could be determined using a score in one of two ways, a highest or a lowest score, as explained in Section VIII.A.4.c), above. See EX1021, 7:57-67 ("'Score' is a numerical evaluation of how well a given hypothesis matches some set of observations. Depending on the conventions in a particular implementation, better matches might be represented by higher scores

(such as with probabilities or logarithms of probabilities) or by lower scores (such as with negative log probabilities or spectral distances"). The Bazzi-Epstein combination therefore renders obvious selecting a sentence candidate interpretation having a highest assigned score as a probable interpretation of an input utterance.

I, Paul Jacobs, do hereby declare and state, 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 further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: 7/21, 2025

Paul Jacobs

Paul S. Jacobs

Paul@jaketech.com
mobile: (703) 307-0467

4105 Faith Ct.
Alexandria VA 22311

Education

Ph.D. in Computer Science	University of California at Berkeley (1985)
S.M. in Applied Mathematics	Harvard University (1981)
A.B. in Applied Mathematics	Harvard University (1981)

Work History

***Founder and President**, Jake Technologies, Inc.* (7/02 -)

- Founded technology services company focused on strategic technology development, product evaluation, and advising corporate counsel and law firms on issues related to intellectual property. Serve as expert consultant in patent litigation, research, and related matters.
- Appointed by the office of the Secretary of Commerce to the Patent Public Advisory Committee, advising the Under Secretary and Director of the U.S.. Patent and Trademark Office from 2012 – 2015.

***Chief Technology Officer**, Primus Knowledge Solutions, Inc.* (6/01 – 7/02)

- Played senior executive role in integration of AnswerLogic into Primus, a publicly-held CRM software company focusing on knowledge capture and delivery in support center and web self-service applications.
- As CTO, responsible for technology "evangelism" both inside of and outside of Primus, including articulating company strategy and vision to analysts, customers, and prospects and working with product and product management teams.

***President, Chief Operating Officer, and Chief Technology Officer**, AnswerLogic, Inc.* (10/99 – 5/01)

- Joined start-up at close of Series A financing, led through Series B financing and initial product launch and made sale of company to Primus Knowledge Solutions.
- Led technology and general management of over 90 employees, including over 50 linguists and over 30 engineers. Only senior executive retained after Primus acquired AnswerLogic.

***Managing Vice President, Electronic Commerce**, SRA International* (9/98 – 10/99)

- Managed NetOwl product line and related IsoQuest technologies for parent company; SRA derived over $10 million in product licenses and millions more in associated services from NetOwl and related products, with over 100 customers. SRA, which soon went public, was ranked in Business Week's Top 20 Private IT firms in 1999, cited for e-mail software based on NetOwl.

***President and Chief Executive Officer**, IsoQuest, Inc.* (7/96 – 9/98)

- Founded and led advanced Internet-focused knowledge management software company, backed by SRA International. IsoQuest's main product, NetOwl, became the leading engine for automatic information extraction as well as forming the basis for the parent company's product offering. In 1998, with IsoQuest at break-even, SRA re-integrated the subsidiary for strategic reasons.
- Responsible for all major financial, strategy, staffing and management decisions
- Conceived product strategy, launched and sold over $2 million of new product (NetOwl) to 20 clients in first two years; brought company to break-even.

- Personally guided marketing, sales, and product management until executives in each area were hired.
- Landed major accounts, including Thomson Corporation, Knight Ridder, Bloomberg, Lexis/Nexis, Disclosure, NewsEdge, and Infoseek.

***Director of Product Marketing**, SRA International* (7/95 – 6/96)

- Led launch of ($200M) consulting firm's first software product (NameTag)
- Developed strategic marketing, branding, product management, and marketing communications programs, hired and managed agencies and contractors and led product marketing initiatives
- Successfully presented business plan and laid groundwork for formation of new company (IsoQuest)

***Director of Media Information Technologies**, SRA International* (8/94 – 6/95)

- Developed and implemented long-range technology transfer plan
- Supported SRA business units (media, intelligence, others) in new initiatives and major sales opportunities
- Led search engine re-design and deployment for new venture (Picture Network International)

***Computer Scientist**, GE Corporate Research and Development* (8/85-8/94)

- Led GE's initiatives in natural language text processing. Built world's leading team in automatic information extraction from free text. Received numerous management awards. Whitney Gallery of Technical Achievers
- Principal investigator for four years in U.S. government's Tipster program (Tipster was cited with the "Hammer" award as a national reinvention laboratory)
- Proposed, won, and deployed advanced text management projects in GE Information Services, Aircraft Engines, Aerospace, Capital Services, and external contracts with ARPA and other agencies.
- Completed GE Technical Management Course
- Spent one year as Visiting Associate Professor at University of Pennsylvania (Computer and Information Sciences, Institute for Research in Cognitive Science)
- Consulted with Infonautics, Inc. (start-up in Wayne, PA) – helped map out initial business plan and technical approach to Homework Helper, their first product.

Professional Activities

Author of over 50 published journal articles and technical papers. Holder of two U.S. Patents. Author of *Text-Based Intelligent Systems* (Lawrence Erlbaum Associates, 1992).
Adjunct Professor, College of Information Studies (iSchool), University of Maryland. Currently teach Information Architecture. Delivered 2008 commencement address.
Member of Patent Public Advisory Committee (PPAC), US Dept. of Commerce, 2012-2015
Member of USACM (Public Policy Committee of Major Computer Trade Association)
Member of Intellectual Property Committee of the USACM
Executive Committee, Association for Computational Linguistics, 1997-9

Other Activities

Accomplished skier, distance runner. Iron Man Triathlete. Winner, Golden Gate Marathon, 1983, 1984. Enjoy exotic travel, adventure sports, hiking with my family.

APPENDIX A

PAUL S. JACOBS
PUBLICATIONS, PROFESSIONAL AND SCHOLARLY ACTIVITIES

Refereed Journal Articles

P. Jacobs. PHRED: A generator for natural language interfaces. *Computational Linguistics 11 (4),* pp. 219-242, 1985. Also appearing in L. Bolc and D. McDonald (eds.), *Natural Language Generation Systems*, Springer Verlag, 1988

P. Jacobs. Knowledge intensive natural language generation. *Artificial Intelligence 33 (3),* pp. 325-378, 1987.

L. Rau and P. Jacobs. NL ∩ IR: Natural language for information retrieval. *International Journal of Intelligent Systems 4 (3),* pp. 319-343, 1989.

L. Rau, P. Jacobs, and U. Zernik. Information extraction and text summarization using linguistic knowledge acquisition. *Information Processing and Management 25 (4),* pp. 419-428, 1989.

P. Jacobs and L. Rau. SCISOR: A system for extracting information from on-line news. In *Communications of the Association for Computing Machinery 33 (11),* pp. 88-97, November 1990.

P. Jacobs. TRUMP: A transportable language understanding program. *International Journal of Intelligent Systems 7 (3),* pp. 245-276, 1992.

P. Jacobs and L. Rau. Innovations in text interpretation. *Artificial Intelligence 63 (1-2),* pp. 143-191, 1993.

P. Jacobs. Using statistical methods to improve knowledge-based news categorization. *IEEE Expert, 8(2),* pp. 13-23, Apr. 1993.

Refereed Conference Proceedings

P. Jacobs. Generation in a natural language interface. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, Germany, 1983.

P. Jacobs and L. Rau. Ace: Associating language with meaning. In *Proceedings of the Sixth European Conference on Artificial Intelligence*, Pisa, Italy, 1984. Reprinted in T. O'Shea (ed.), *Advances in Artificial Intelligence*, North Holland, 1986.

P. Jacobs. The KING natural language generator. In *Proceedings of the Seventh European Conference on Artificial Intelligence, Brighton, England*, 1986. Reprinted in B. du Boulay, D. Hogg, and L. Steels (eds.), *Advances in Artificial Intelligence II*, North Holland, 1987.

P. Jacobs. Knowledge structures for natural language generation. In *Proceedings of the 11th International Conference on Computational Linguistics*, pp. 554-559, Bonn, 1986.

P. Jacobs. Language analysis in not-so-limited domains. In *Proceedings of the Fall Joint Computer Conference*, Dallas, 1986.

P. Jacobs. A knowledge framework for natural language analysis. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, Milan, 1987.

D. Besemer and P. Jacobs. FLUSH: A flexible lexicon design. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pp. 186-193, Palo Alto, 1987.

P. Jacobs. Achieving bidirectionality. In *Proceedings of the 12th International Conference on Computational Linguistics*, pp. 267-274, Budapest, 1988.

P. Jacobs. Concretion: Assumption-based understanding. In *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, 1988.

P. Jacobs. Why text planning isn't planning. In *Proceedings of the AAAI Workshop on Text Planning and Generation*, St. Paul, Minnesota, 1988.

P. Jacobs and U. Zernik. Acquiring lexical knowledge from text: A case study. In *Proceedings of the National Conference on Artificial Intelligence*, St. Paul, Minnesota, 1988.

P. Jacobs and L. Rau. Natural language techniques for intelligent information retrieval. In *Proceedings of the 11th International Conference on Research and Development in Information Retrieval*, Grenoble, France, 1988.

P. Jacobs. Extensible natural language interfaces. In *Proceedings of the 17th ASIS Mid-Year Meeting*, Ann Arbor, 1988.

P. Jacobs and L. Rau. A friendly merger of conceptual expectations and linguistic analysis in a text processing system. In *Proceedings of the Fourth IEEE Conference on Artificial Intelligence Applications*, pp. 351-356, San Diego, 1988.

L. Rau and P. Jacobs. Integrating top-down and bottom-up strategies in a text processing system. In *Proceedings of the Second Conference on Applied Natural Language Processing (ANLP)*, pp. 129-135, Austin, 1988.

P. Jacobs and L. Rau. The GE NLToolset: A software foundation for intelligent text processing. In *Proceedings of the 13th International Conference on Computational Linguistics*, Helsinki, 1990.

U. Zernik and P. Jacobs. Tagging for learning: Collecting thematic relations from Corpus. In *Proceedings of the 13th International Conference on Computational Linguistics*, pp. 34-39, Helsinki, 1990.

P. Jacobs, G. Krupka, L. Rau, N. Sondheimer, and U. Zernik. Generic text processing. In *Proceedings of the Third DARPA Speech and Natural Language Workshop*, 359-364, Somerset, PA, 1990.

P. Jacobs. To parse or not to parse: Relation-driven text skimming. In *Proceedings of the 13th International Conference on Computational Linguistics*, pp. 194-198, Helsinki, 1990.

L. Rau and P. Jacobs. Creating segmented databases from free text for text retrieval. In *Proceedings of the 14th International Conference on Research and Development in Information Retrieval*, pp. 337-346, Chicago, 1991.

P. Jacobs, G. Krupka, and L. Rau. Lexico-semantic pattern matching as a companion to parsing in text understanding. In *Proceedings of the Fourth DARPA Speech and Natural Language Workshop*, Asilomar, CA, 1991.

P. Jacobs. Parsing run amok: Relation-driven control for text analysis. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pp. 315-321, San Jose, 1992.

P. Jacobs. Joining statistics with NLP for text categorization. In *Proceedings of the Third Conference on Applied Natural Language Processing*, pp. 178-185, Trento, Italy, 1992.

P. Jacobs, G. Krupka, and L. Rau. A Boolean approximation method for query construction and topic assignment in TREC. In *Proceedings of the Second Annual Symposium on Document Analysis and Retrieval*, Las Vegas, 1993.

P. Jacobs. Word sense acquisition for multilingual text interpretation. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)*, Vol. II, pp. 665-671, 1994.

P. Jacobs. Browsing vs. Surfing: The Next Generation of Search and Retrieval. In *Proceedings of the National On-Line Conference*. New York, 1998.

Books and Book Chapters

P. Jacobs. KING: A knowledge-intensive natural language generator. In G. Kempen (ed.), *Natural Language Generation: Recent Advances in Artificial Intelligence, Psychology, and Linguistics*, Kluwer Academic Publishers, 1987.

P. Jacobs. Two hurdles for natural language systems. In R. Freedle, (ed.), *Artificial Intelligence and the Future of Testing*, pp. 257-277, Lawrence Erlbaum Associates, Hillsdale, NJ, 1990.

P. Jacobs. Making sense of lexical acquisition. In U. Zernik, (ed.), *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon*, pp. 29-44, Lawrence Erlbaum Associates, Hillsdale, NJ, 1991.

P. Jacobs. Integrating language and meaning in structured inheritance networks. In J. Sowa, (ed.), *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, pp. 527-542, Morgan Kaufman Publishers, San Mateo, CA, 1991.

P. Jacobs. Text power and intelligent systems. In *Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval*, pp. 1-8, Lawrence Erlbaum Associates, Hillsdale, NJ, 1992.

P. Jacobs (ed). *Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1992.

P. Jacobs. Why Text Planning (Still) Isn't Planning. In A. Ortony, J. Slack, and O. Slack (eds.), *Communication from an Artificial Intelligence Perspective*, Springer Verlag, Heidelberg, 1992.

P. Jacobs. Information extraction. In R. Cole, J. Mariani, H. Uszkoreit, A. Zaenen, and V. Zue (eds.), *Survey of the State of the Art in Human Language Technology*, pp. 230-232, Cambridge University Press, 1997.

Other Published Articles

P. Jacobs. Generation in the UNIX Consultant System, in D. Roesner (ed.), *International Workshop on Language Generation*, Institut fuer Informatik, Universitaet Stuttgart, 1983.

P. Jacobs. Teaching language to computers. *IEEE Potentials*, October, 1986.

P. Jacobs and L. Rau. Software for intelligent text processing. In *Proceedings of the 13th International Conference on Computational Linguistics (project notes)*, Helsinki, Vol. 3, 1990, pp 373-375.

G. Krupka, L. Iwanska, P. Jacobs, and L. Rau. GE NLToolset test results and analysis. In *Proceedings of the Third Message Understanding Conference (MUC-3)*, San Diego, 1991.

G. Krupka, P. Jacobs, L. Rau and L. Iwanska. Description of the GE NLToolset system as used in MUC-3. In *Proceedings of the Third Message Understanding Conference (MUC-3)*, San Diego, 1991.

P. Jacobs, G. Krupka, and L. Rau. The right tools for the job: Multiple language analysis strategies in the GE NLToolset. In *Proceedings of the Workshop on Fully Implemented Natural Language Systems*, Trento, Italy, March, 1992.

P. Jacobs, G. Krupka, L. Rau, T. Kaufmann and M. Mauldin. GE-CMU: Description of the TIPSTER/SHOGUN system as used for MUC-4. In *Proceedings of the Fourth Message Understanding Conference (MUC-4)*, McLean, VA, 1992.

L. Rau, G. Krupka, P. Jacobs, I. Sider and L. Childs. GE NLToolset test results and analysis. In *Proceedings of the Fourth Message Understanding Conference (MUC-4)*, McLean, VA, 1992.

P. Jacobs, G. Krupka, L. Rau, M. Mauldin, T. Mitamura, T. Kitani, I. Sider, and L. Childs. GE-CMU: description of the SHOGUN system used for MUC-5. In *Proceedings of the Fifth Message Understanding Conference (MUC-5)*, pp. 109-120, 1993.

P. Jacobs, G. Krupka, L. Rau, M. Mauldin, T. Mitamura, T. Kitani, I. Sider, and L. Childs. The TIPSTER/SHOGUN Project. In *Proceedings of the TIPSTER Text Phase I Final Meeting*, pp. 209-221, Morgan Kaufmann, 1993.

P. Jacobs. GE in TREC-2: Results of a Boolean approximation method for routing and retrieval. In *Proceedings of the Second Text Retrieval Conference (TREC-2)*, D. Harman (ed.), pp. 191-199, NIST Special Publication 500-215, 1994.

P. Jacobs. Natural language processing: A brief history for skeptics. *ServerWorld*, February 2001.

Technical Reports Not Published Elsewhere

P. Jacobs. *A Knowledge-Based Approach to Language Production*. Ph. D. Thesis, University of California, Berkeley. Available as Computer Science Division Technical Report 86/254, 1985.

P. Jacobs. A cooperative natural language directory system. General Electric Corporate Research and Development, Report 88CRD025, 1988.

Research Grants

Principal investigator of Tipster Phase I (GE, 1991, $2.8 million) and Phase II (GE, 1993)
$10s of millions in commercial financing for advanced research and development from 1994 through 2000.

Professional Societies and Committees

Member of IEEE, ACM, AAAI, and ACL for many years.
Program committees:
- AAAI National Conference on Artificial Intelligence
- ACL Annual Conference
- ACL Conference on Applied Natural Language Processing
- ACM SIGIR Conference on Research and Development in Information Retrieval
- ARPA Speech and Natural Language Conference
- Others

Program chair, AAAI Symposium on Text-Based Intelligent Systems, 1990 (led to the publication of the book, *Text-Based Intelligent Systems* and to an increase in cross-disciplinary work in computational linguistics and information retrieval)
Program chair, ACL Conference on Applied National Language Processing, Stuttgart, Germany, 1994.
Executive Committee of the ACL, 1995-1998
USACM Public Policy Committee, Intellectual Property Subcommittee, 2005 -

U.S. Patents

5,243,520 Sense Discrimination System and Method, Sept. 7, 1993
5,251,129 Method for Automated Morphological Analysis of Word Structure, Oct. 5, 1993