Journal of Experimental and Theoretical Artificial Intelligence, 10, 487-493, 1998.

Artificial Intelligence and Other Approaches to Speech Understanding: Reflections on Methodology

Nigel Ward

University of Tokyo

Abstract

This paper characterizes the methodology of Artificial Intelligence by looking at research in speech understanding, a field where AI approaches contrast starkly with the alternatives, particularly engineering approaches. Four values of AI stand out as influential: ambitious goals, introspective plausibility, computational elegance, and wide significance. The paper also discusses the utility and larger significance of these values.

1 Introduction

AI is often defined in terms of the problems it studies. But in fact, AI is not *the* study of intelligent behavior etc., it is *a* way to study it. This is evident from the way AI is done in practice. This paper illustrates this by contrasting AI and alternative approaches to speech understanding. By so doing it brings out some key characteristics of AI methodology.

This paper is primarily written for a general AI audience interested in methodological issues, complementing previous work (Cohen 1991; Brooks 1991a). It is also written for any AI researchers who are contemplating starting a project in speech understanding — it is intended to be the paper that, if available earlier, might have saved me four years of wasted effort. This paper does not provide technical discussion of specific issues in speech understanding; for that, see (Ward 1996) and the references given below.

The points made here are based on, first, a review of the full spectrum of published work on speech understanding, including AI work from the 1970s to the present (Reddy *et al.*

[0] Acknowledgements: Thanks to Dan Jurafsky, Jane Edwards, and Wataru Tsukahara for comments, and to the Sound Technology Promotion Foundation for support.

1973; Woods & Makhoul 1973; Klatt 1977; Erman *et al.* 1980; Woods 1980; Hayes *et al.* 1987; Kitano *et al.* 1989; Baggia & Rullent 1993; Nagao *et al.* 1993; Kawahara *et al.* 1994; Hauenstein & Weber 1994; Cochard & Oppizzi 1995; Weber & Wermter 1996), and also engineering work, including (Moore *et al.* 1989; Lee 1994; Nguyen *et al.* 1994; Moore 1994; Alshawi & Carter 1994; Hirschman 1994; Levin & Pieraccini 1995; Jurafsky *et al.* 1995; Seneff 1995; Moore *et al.* 1995; Pallett & Fiscus 1995) and psycholinguistic work, including (McClelland 1987; Norris 1993; Nygaard & Pisoni 1995; Cutler 1995); and second, my own experience building an AI speech understanding system (Ward 1992; Ward 1993; Ward 1994a; Ward 1994b; Ward 1995). The discussion is general, abstract, and simplistic: no specific project is characterized accurately, there is no attempt to distinguish among the various AI approaches, and there is no discussion of approaches part way along the continuums between AI and rival approaches. The aim is to bring together some diverse observations; no individual point is claimed to be original.

2 Characteristics of the AI Approach

This section discusses the AI approach in terms of four key values of AI: ambitious goals, introspective plausibility, computational elegance, and wide significance. These remarks apply specifically to classical AI in its pure form, but are also more generally relevant, as discussed in §4.

2.1 Ambitious Goals and Bold Leaps

AI speech research, like AI research more generally, is motivated not by what can be achieved soon, but by a long-term vision. For speech, this has been the idea of a system able to produce an optimal interpretation based on exhaustive processing of the input.

Engineers prefer to set goals towards which progress can be measured.

AI speech research tends, since existing systems are nowhere near optimal, to seek breakthroughs and radically new perspectives.

Engineers tend to proceed by improving existing systems.

AI speech research, like other AI research, tries to solve problems in their most general form, often by trying to construct a single, general-purpose system.

Engineering is in large part the creation of solutions for specific problems, and the resulting accumulation of know-how useful for solving other problems.

2.2 Introspective Plausibility

AI speech research, like AI more generally, often makes recourse to introspection about human abilities. This subsection illustrates the use of introspection at four phases: when setting long-term goals, when setting short-term goals, for design, and when debugging.

1. When setting long-term research goals, AI speech research, like AI research in general, aims to duplicate apparent human abilities. In particular, many AI speech researchers are inspired by the apparent near perfection of human speech understanding: the fact that people can understand just about anything you say to them, and even repeat back to you the words you said.

Scientific approaches to human speech understanding, in contrast, find it more productive to focus on what can be learned from the limitations of human speech understanding. (A simple demonstration of these limitations can be had by simply recording a conversation and later listening to it repeatedly; you will discover that you missed a lot when hearing it live.) In general, the feeling that perception and understanding is complete is an illusion of introspection (Brooks 1991a; Dennett 1991).

2. When setting short-term goals, AI speech research, like AI more generally, values systems which do things which seem, introspectively, clever. Such cleverness is often found not in the overall performance but in the details of processing in specific cases. For speech understanding, AI has emphasized the use of reasoning to compensate for failures of the low-level recognition component, often by selecting or creating word hypotheses for which the recognizer had little or no bottom-up evidence. Doing this can involve arbitrarily obscure knowledge and arbitrarily clever inferences, which makes for impressive traces of system operation.

Engineers typically design and tune systems to work well on average; they seldom show off cleverness in specific cases. Few engineered speech systems do much explicit reasoning, and none bother to explicitly correct mis-recognitions — rather, they simply barge on to compute the most likely semantic interpretations.

3. For design, AI speech research, like AI more generally, takes inspiration from human "cognitive architecture", as revealed by introspection. For speech, this has led to the use of protocol studies, done with a committee or with one man and a scratchpad. Both suggest

a view of speech understanding as problem solving, and suggest a process where diverse knowledge sources cooperate by taking turns — for example, with a partially recognized input leading to a partial understanding, that understanding being used to "figure out" more words, leading to a better recognition result, then a better understanding, and so on. This has been fleshed out into, for example, "blackboard models", which include a focus on explicit representation of hypotheses, a focus on the interaction of knowledge sources, a focus on the scheduling of processing, and an image of iterative refinement over many cycles of computation.

Scientific approaches, on the other hand, do not consider introspection to be reliable or complete. For example, introspection can easily focus on conscious "figuring out", explicit decisions, and the role of grammar, but probably not on automatic processing, parallel processing of multitudes of hypotheses, and the role of prosody.

4. For development, AI speech research, like AI more generally, values systems which people can understand intuitively. This makes it possible to debug by examining behavior on specific cases and adjusting the system until it works in a way that introspectively seems right. By doing so the developer can see if the system is a proper implementation of his vision. More important, he can get a feel for whether that vision is workable. In other words, the developer can use experience with the internal workings of a program to leverage introspection about how a cognitive process might operate. This technique is perhaps the most distinctive aspect of AI methodology; it could be called "the hacker's path to intelligent systems" or perhaps "computational introspection".

Engineers focus on the desired input-output behavior of a system and design algorithms to achieve it. They typically do not care about the introspective plausibility of the intermediate steps of the computation.

2.3 Computational Elegance

AI speech research, like AI research more generally, postulates that knowledge is good and that more knowledge is better. In order to bring more knowledge to bear on specific decisions, integration of knowledge sources is considered essential. For speech, this means, most typically, wanting to use the full inventory of higher-level knowledge, including knowledge of syntax, semantics, domain, task and current dialog state, at the earliest stages of recognition and understanding.

Engineers focus on performance, for which more knowledge may or may not be worthwhile, especially when processing time and memory are finite. Whether a specific type of knowledge is worthwhile, and if so when to apply it, are determined by experiment.

AI speech research, like other AI research, involves an aesthetic sense of what designs are good. In particular, to build systems that can be extended and scaled up, AI researchers generally feel that elegant designs are required. For speech, the meaning of "elegant" has changed along with broader fashions in computer science — at various times it has included: explicit (declarative) representation of hypotheses and control structures, emergent properties, multiprocessor (distributed) parallelism, fine-grained (connectionist, massive) parallelism, uniformity, heterogeneity, symbolic reasoning, evidential reasoning, and so on.

Engineers prefer to actually try to build big systems, rather than just build prototypes and argue that they will scale up.

2.4 Wide Significance

AI speech research, like AI research in general, has a larger purpose: researchers don't just want to solve the problem at hand, they also want their solution to inspire other work.

AI speech researchers have generally wanted their work to be relevant to other problems in natural language processing and to linguistics, and have sought out and focused on phenomena of interest to those fields, such as ambiguity.

AI speech researchers have also wanted to be relevant for the larger AI community. They have emphasized analogies relating speech understanding to other topics, such as search and planning. They have also emphasized ties to general AI issues, such as the Attention Problem, the Constraint Propagation Problem, the Context Problem, the Disambiguation Problem, the Evidential Reasoning Problem, the Knowledge Integration Problem, the Knowledge Representation Problem, the Noisy Input Problem, the Real-World Problem, the Reasoning with Uncertainty Problem, the Sensor Fusion Problem, the Signal-to-Symbol Problem, and a few others.

AI speech researchers have also tried to make their work relevant for computer science more generally. Based on insights from speech understanding they have called for new architectures for computer networks, for software systems, and for computer hardware.

Engineers prefer to work on goals rather than "interesting" problems. They also prefer to work on speech for its own sake, rather than for what it reveals about other problems.

They tend to think in terms of real problems, not abstract ones.

3 Outcome

For speech understanding, and more generally, AI approaches have seemed more promising than traditional science and engineering. This is probably because AI methodology exploits introspection (§2.2) and aesthetic considerations (§2.3), both of which seem to provide the right answers with relatively little effort.

However AI methodology has not fulfilled this promise for speech. The engineering approach, in contrast, has produced solid and impressive results.

The significance of this is not obvious. Some AI researchers believe that this means that the speech community should be "welcomed back" to AI, as argued by several recent editorials. But the values of AI and mainstream (engineering-style) speech research are so different, as seen in §2, that reconciliation does not seem likely.

Other AI researchers take the view that the results of engineering work are not interesting; presumably meaning that they are compelling neither introspectively or aesthetically. Many further believe that the AI approach to speech will be vindicated in the end. A few strive towards this goal (I was one). However, AI goals conflict with other more important goals, and so it is hard to be optimistic about future attempts to apply AI methodology to the speech understanding problem.

The conclusions of this case study are thus in line with the conclusions of Brooks' case study (Brooks 1991a). For both speech understanding and robotics, AI methodology turns out to be of little value. Whether this is also true in other cases is a question of great interest.

4 Larger Significance

The values of AI outlined in §2 actually best characterize the methodology of classical AI, ascendent in the 1970s, but less popular in recent years, with the proliferation of variant approaches to AI, including some, such as connectionism and Brooks-style[1] AI (Brooks 1991a; Brooks 1991b), which reject many of the values of mainstream AI. Nevertheless, classical AI methodology still has wide influence.

[1] It is interesting to note that adopting a Brooks-style approach to speech leads to the view that, for building spoken dialog systems, it is not speech *understanding* that is the key problem (Ward 1997).

For one thing, the values of §2 were important when the topics and tasks of AI were being established. As a result, those working on knowledge representation, planning, reasoning, distributed intelligence, user modeling, etc. today, even if they do not share the classical values of AI, are working on problems posed by researchers who did.

Moreover, despite changes in AI methodology (mostly in the direction of "methodological respectability", in the sense of importing values and techniques from engineering and science), the values of §2 are not only still alive, but remain intrinsic to the existence of AI as a field. It is true that, at recent AI conferences, most papers do not explicitly refer to these values, but the fact that the papers appear at AI conferences at all is tribute to them. If AI researchers did not rely on introspection, have grand goals, or value aesthetic considerations, most of them would drift off to conferences on computational psychology, user interface engineering, specific applications, etc. And without the belief that results should have wide significance and utility, groups of AI researchers would drift off to separate fields of speech, vision, motion, etc.

In any case, it is clear that AI is a distinctive field of study in many ways. This paper has been an endeavor to pinpoint just what it is that is special about AI.

References

Alshawi, Hiyan & David Carter (1994). Training and Scaling Preference Functions for Disambiguation. *Computational Linguistics*, 20:635–448.

Baggia, Paolo & Claudio Rullent (1993). Partial Parsing as a Robust Parsing Strategy. In *1993 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. II–123–126.

Brooks, Rodney A. (1991a). Intelligence Without Reason. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, pp. 569–595.

Brooks, Rodney A. (1991b). Intelligence Without Representation. *Artificial Intelligence*, 47:139–159.

Cochard, Jean-Luc & Olivier Oppizzi (1995). Reliability in a Multi-agent Spoken Language Recognition System. In *4th European Conference on Speech Communication and Technology (Eurospeech'95)*, pp. 75–78.

Cohen, Paul R. (1991). A Survey of the Eighth National Conference on Artificial Intelligence. *AI Magazine*, 12:16–41.

Cutler, Anne (1995). Spoken Word Recognition and Production. In Joanne L. Miller & Peter D. Eimas, editors, *Speech, Languge, and Communication*. Academic Press.

Dennett, Daniel C. (1991). *Consciousness Explained*. Penguin.

Erman, Lee D., Frederick Hayes-Roth, Victor R. Lesser, & D. Raj Reddy (1980). The Hearsay-II Speech-Understanding System: Integrating Knowledge to Resolve Uncertainty. *Computing Surveys*, 12:213–253.

Hauenstein, A. & H. Weber (1994). An Investigation of Tightly Coupled Time Synchronous Speech Language Interfaces Using a Unification Grammar. In *AAAI Workshop on the Integration of Natural Language and Speech Processing*, pp. 42–49.

Hayes, Phillip J., Alexander G. Hauptmann, Jaime G. Carbonell, & Masaru Tomita (1987). Parsing Spoken Language: a Semantic Caseframe Approach. Technical Report CMU-CMT-87-103, Carnegie Mellon University, Center for Machine Translation. expanded version of a paper in COLING86.

Hirschman, Lynette (1994). The Roles of Language Processing in a Spoken Language Interface. In David B. Roe & Jay G. Wilpon, editors, *Voice Communication between Humans and Machines*, pp. 217–237. National Academy Press.

Jurafsky, Daniel *et al.* (1995). Using a Stochastic Context-free Grammar as a Language Model for Speech Recognition. In *1995 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 180–192.

Kawahara, Tatsuya, Masahiro Araki, & Shuji Doshita (1994). Heuristic Search Integrating Syntactic, Semantic and Dialog-level Constraints. In *1994 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. II–25–28.

Kitano, Hiroaki, Hideto Tomaebechi, Teruko Mitamura, & Hitoshi Iida (1989). A Massively Parallel Model of Speech-to-Speech Dialog Translation. In *European Conference on Speech Communication and Technology (Eurospeech'89)*.

Klatt, Dennis H. (1977). Review of the ARPA Speech Understanding Project. *Journal of the Acoustical Society of America*, 62:1324–1366. reprinted in *Readings in Speech Recognition*, Alex Waibel and Kai-Fu Lee, eds., Morgan Kaufmann, 1990.

Lee, Chin-Hui (1994). Stochastic Modeling in Spoken Dialogue System Design. *Speech Communication*, 15:311–322.

Levin, Esther & Roberto Pieraccini (1995). Concept-Based Spontaneous Speech Understanding System. In *4th European Conference on Speech Communication and Technology (Eurospeech'95)*, pp. 555–558.

McClelland, James L. (1987). The Case for Interationism in Language Processing. In Max Coltheart, editor, *Attention and Performance XII: The Psychology of Reading*, pp. 3–36. Lawrence Erlbaum Associates.

Moore, Robert, Fernando Pereira, & Hy Murveit (1989). Integrating Speech and Natural Language Processing. In *Speech and Natural Language Workshop*, pp. 243–247. Morgan Kaufmann.

Moore, Robert C. (1994). Integration of Speech with Natural Language Understanding. In David B. Roe & Jay G. Wilpon, editors, *Voice Communication Between Humans and Machines*, pp. 254–271. National Academy Press.

Moore, Robert C., Douglas Appelt, John Dowding, J. Mark Gawron, & Douglas Moran (1995). Combining Linguistic and Statistical Knowledge Sources in Natural-Language Processing for ATIS. In *Proceedings of the Spoken Language Systems Technology Workshop*, pp. 261–264. Morgan Kaufmann.

Nagao, Katashi, Koiti Hasida, & Takashi Miyata (1993). Understanding Spoken Natural Language with Omni-Directional Information Flow. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pp. 1268–1274.

Nguyen, Long, Richard Schwartz, Ying Zhao, & George Zavaliagkos (1994). Is N-Best Dead? In *Proceedings of the Human Language Technology Workshop*, pp. 411–414. Morgan Kaufmann.

Norris, Dennis (1993). Bottom-Up Connectionist Models of 'Interaction'. In Gerry Altman & Richard Shillcock, editors, *Cognitive Models of Speech Processing*, pp. 211–234. Lawrence Erlbaum Associates.

Nygaard, Lynne C. & David B. Pisoni (1995). Speech Perception: New Directions in Research and Theory. In Joanne L. Miller & Peter D. Eimas, editors, *Speech, Languge, and Communication*. Academic Press.

Pallett, David S. & Jonathan G. Fiscus (1995). 1994 Benchmark Tests for the ARPA Spoken Language Program. In *Proceedings of the Spoken Language Systems Technology Workshop*, pp. 5–36. Morgan Kaufmann.

Reddy, D. Raj, Lee D. Erman, & Richard B. Neely (1973). A Model and a System for Machine Recognition of Speech. *IEEE Transactions on Audio and Electroacoustics*, 21:229–238. reprinted in *Automatic Speech and Speaker Recognition*, N. Rex Dixon and Thomas B. Martin (eds), IEEE Press, 1979, pages 272–281.

Seneff, Stephanie (1995). Integrating Natural Language into the Word Graph Search for Simultaneous Speech Recognition and Understanding. In *4th European Conference on Speech Communication and Technology (Eurospeech'95)*, pp. 1781–1784.

Ward, Nigel (1992). An Evidential Model of Syntax for Understanding. Technical Report 88-3, Information Processing Society of Japan, Natural Language Working Group, Tokyo.

Ward, Nigel (1993). On the Role of Syntax in Speech Understanding. In *Proceedings of the International Workshop on Speech Processing*, pp. 7–12.

Ward, Nigel (1994a). An Approach to Tightly-Coupled Syntactic/Semantic Processing for Speech Understanding. In *AAAI Workshop on the Integration of Natural Language and Speech Processing*, pp. 50–57. ftp: ftp.sanpo.t.u-tokyo.ac.jp/pub/nigel/papers/integration94.ps.Z.

Ward, Nigel (1994b). A Lightweight Parser for Speech Understanding. In *International Conference on Spoken Language Processing*, pp. 783–786.

Ward, Nigel (1995). The Spoken Language Understanding Mini-challenge. ftp: ftp.sanpo.t.u-tokyo.ac.jp/pub/nigel/lotec2-slum. (corpus and evaluation software).

Ward, Nigel (1996). Second Thoughts on an Artificial Intelligence Approach to Speech Understanding. In *14th Spoken Language and Discourse Workshop Notes (SIG-SLUD-14)*, pp. 16–23. Japan Society for Artificial Intelligence. ftp: ftp.sanpo.t.u-tokyo.ac.jp/pub/nigel/papers/second-thoughts.ps.Z.

Ward, Nigel (1997). Responsiveness in Dialog and Priorities for Language Research. *Systems and Cybernetics*, 28(6):521–533.

Weber, Volker & Stefan Wermter (1996). Using Hybrid Connectionist Learning for Speech/Language Analysis. In Stefan Wermter, Ellen Riloff, & Gabriele Scheler, editors, *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*. Springer.

Woods, W. A. (1980). Control of Syntax and Semantics in Continuous Speech Understanding. In *Spoken Language Generation and Understanding*, pp. 337–364. D. Reidel.

Woods, W. A. & J. Makhoul (1973). Mechanical Inference Problems in Continuous Speech Understanding. In *Proceedings of the Third International Joint Conference on Artificial Intelligence*, pp. 200–207. revised version appears in *Artificial Intelligence*, 5(1), 1974, pp 73–91.