ELSEVIER

Speech Communication 31 (2000) 293–307

SPEECH COMMUNICATION

www.elsevier.nl/locate/specom

Is ASR ready for wireless primetime: Measuring the core technology for selected applications

Harry M. Chang

SBC Technology Resources, Speech & Computing Technologies, 9505 Arboretum Blvd., Austin, TX 78759, USA

Received 18 March 1999; received in revised form 22 June 1999; accepted 25 November 1999

Abstract

It is estimated that by the end of 2001 as many as 500 million people worldwide will use cellular services. The nature of hands-busy and eyes-busy situations inherent in the anywhere and anytime wireless communication paradigm presents exciting marketing opportunities and, at the same time, unique technical challenges to the current-generation ASR technology and their new applications. Current industry trends clearly show that incorporating ASR technology into existing or new wireless services as a replacement for touch-tone input is a natural progression in user interface. But is the current-generation ASR technology ready for prime time over wireless channels? Both qualitative and quantitative assessments for the core technology must be adopted by the industry before answering this question. In this paper, we will describe a set of benchmark tasks designed to evaluate the state-of-the-art ASR technologies from a wireless perspective and present the results of these benchmark tests on two commercially available software-based ASR systems that represent the best core ASR technology on the market. © 2000 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Es wird geschätzt, daß Ende 2001 ebensoviel als 500 Million Leute weltweit zellulare Dienstleistungen verwendet. Die Natur der Hand-besetzten und Auge-besetzten Situationen zugehörig in überall und immer wenn drahtloses Kommunikation Paradigma aufregende Marketing-Gelegenheiten und gleichzeitig eindeutige technische Herausforderungen der Technologie des Aktuellerzeugung ASR und ihre Neuanmeldungen vorstellt. Aktuelle Industrietendenzen zeigen dem verbindene ASR-Technologie offenbar in die vorhandenen oder neuen drahtlosen Dienstleistungen, da ein Wiedereinbau für touch-tone Input eine natürliche Weiterentwicklung in der Benutzerschnittstelle ist. Aber ist die Technologie des Aktuellerzeugung ASR zur Hauptzeit über drahtlosen Führungen betriebsbereit? müssen qualitative und quantitative Einschätzungen für die Kerntechnologie bei der Industrie angenommen werden, bevor man diese Frage beantwortet. In diesem Papier beschreiben wir ein Set Bankettmarkierungsarbeiten, die entworfen werden, um die state-of-the-art ASR-Technologien von einer drahtlosen Perspektive auszuwerten und die Resultate dieser Prüfstandversuche auf zwei kommerziell darzustellen Nutzen. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Automatic speech recognition; Wireless applications; Speech analysis; Performance assessment; Benchmarking; Digits recognition; Noisy speech recognition

E-mail address: hchang@tri.sbc.com (H.M. Chang).

PII: S0167-6393(99)00063-1

1. Introduction

The rapidly growing wireless telephony marketplace of the past five years creates a compelling need for automatic speech recognition (ASR) enabled voice-activated services. A recent survey indicates that a considerable percentage of wireless users (32%) want some type of hands-free voice activation for any services they might want to access from their cell phones (Rietman, 1997). This is quite understandable, given the fact that 83% of all cellular calls in the US are placed inside a vehicle (Rietman, 1997). For this very reason, basic voice-activated dialing (VAD) has become a de facto service pursuit for all major mobile service operators worldwide who try to apply ASR to wireless telephony applications. More recently, ASR has been applied to other wireless applications such as voice messaging and personal access services. The latest deployment of Network Wildfire by Pacific Bell Wireless at their California markets is just one more witness to this long-term trend. As these new applications begin to impact how the mainstream wireless users would use new ASR-based network technologies, it is important to conduct necessary benchmark tests so that we can better predict their field performance. This is one of the major motivations behind this study.

1.1. Technology and system selection

The core ASR technology employed in most wireless applications today can be measured as a function of context free grammars (CFGs) that define exactly which words (or phrases or sentences) are to be recognized. Using these CFGs a series of ASR benchmark tasks can be carefully configured to mirror ASR applications of interest. To be able to handle arbitrary grammar-based recognition tasks on standard Unix systems such as Solaris-based workstations or PCs without using any special purpose hardware plug-in boards was one of the important considerations in our selection process. Other selection criteria included the quality of the R&D programs behind the vendors' products and the representations of the state-of-the-art ASR technologies in the commercial marketplace. Finally, the potential candidates must support telephony applications in both landline and cellular environment.

In this study, we selected two hidden Markov models (HMMs) based ASR systems for the reasons described above. Both ASR systems are host-based. The versions of both systems used during this study were based on Solaris. Both systems support file input modes. This important system feature made it possible to utilize pre-recorded speech databases constructed specifically to characterize the real-world wireless environments such as handset variations, in-vehicle or out-vehicle, driving speeds, and other speech-bearing user attributes. Furthermore, a file input based test framework ensures reproducible test results and does not introduce any artificial channel elements to the test process.

1.2. Scope and focus

To reflect emerging industry trends towards speaker independent ASR applications over both landline and wireless networks, all benchmark tests described in this paper were based on speaker independent speech recognition. In addition, since the primary objective of this study was to establish a set of performance benchmarks for HMM-based ASR technology in general, no effort was made to statistically compare the two ASR systems tested in terms of their performance. Rather, the emphasis was placed on how they would score as a group on different benchmark tests. Because both systems can run on standard Unix workstations without any special hardware components, memory requirements or CPU load were not our concerns. In other words, accuracy issues were the main focus of the study.

2. Overview of testing framework

Speech recognition over the wireless network introduces many confounding variables over which the robustness of ASR technology has to be measured. As handheld cellular phones (analog or digital) now clearly dominate the landscape of the wireless handset market with ever expanding network coverage, it is increasingly difficult to