Unlocking the Potential of the Spoken Word

Douglas W. Oard
University of Maryland, College Park, MD 20742 USA
oard@umd.edu

The best available evidence suggests that the human brain, and in particular the human facility for language, was already well developed at least 50,000 years ago. For almost all of the time since then, the spoken word provided the only practical way of using language to share our understanding of the world with others. To this day, people find spoken expression and its visual correlates (e.g., gesture and facial expression) to be a fluid and compelling way of communicating. About 5,000 years ago, we see the first indications of the emergence of written language. Writing has important features that the spoken word lacks—among the most important are that writing exhibits a degree of permanence that can help to overcome some limitations of human memory. The effects of writing as an innovation are hard to overstate—it proliferated well beyond mere commercial records to play a multifaceted role in complex forms of social organization. This proliferation inspired other innovations: ways of finding documents again, and ways of writing that conveyed the needed context to a reader. The written word has other attractive qualities as well (e.g., you can read it at your own pace), but permanence, findability, and contextualization are responsible for the foundational role of writing in human civilization.

For the past century and a half, inventors have chipped away at those advantages. The earliest known recording of a human voice was made in 1860 by Edward Lyon Scott's phonautograph, although it was not until Edison's 1877 better known phonograph that the human voice could also be reproduced using technology from the same era. Later technologies, from wire recorders through reel-to-reel tape recording, were widely adopted for commercial purposes. It was, however, introduction of the compact cassette in 1962 that ultimately made sound recording technology robust and affordable. By the end of that decade, ordinary people could record hundreds of hours of speech, for media costs of about a dollar an hour. Today, digitized speech is easily acquired (e.g., using any of the world's 2.5 billion mobile phones), easily transferred over digital networks, and easily stored, all for just few cents per hour. Indeed, it would take just $100 or so of networked disk storage to record everything that you will speak or hear this year.

Digital storage is a great equalizer with regard to permanence—the same infrastructure that can reliably store digital text can equally well store digital speech. Why, then, don't we record our lives in this way? Actually, some people do. For example, researchers at Carnegie Mellon University crafted a memory aid by recording their side of conversations and then using face recognition to cue up audio from an earlier meeting—no more forgetting people's names! Gordon Bell at Microsoft has gone further, assembling digital materials from his entire life. This works well for some things (all your email, for example), but speech is not one of them—searching through large collections of spontaneously produced speech has remained a challenge.

This situation is about to change, however. Commercial “media management” systems can already reliably find specific content in the well articulated speech of news announcers, and laboratory systems are now able to handle much of the substantial variations in speaking styles that have made automatic transcription of interviews, meetings, and telephone conversations difficult. Although hardware costs are higher for speech than for born-digital text (around a factor of 100 for storage, and perhaps a factor of 1,000 for processing), it is possible today to acquire, store and process digitized speech at lower cost than was possible for born-digital text at the dawn of the Web. Robust accommodation to noisy environments and unfamiliar words remain important challenges, however, limiting the tasks for which present speech technology can productively be applied.

As increasingly capable systems emerge from the laboratory, we will soon find ourselves in a world in which speech need no longer be ephemeral. How will that change our society? Certainly no one can know for sure, but it’s not hard to envision some questions that might arise. The Carnegie Mellon system recorded only one side of the conversation because it is illegal in Pennsylvania (and eleven other states) to record full conversations without the explicit consent of all parties. Will a new balance between social costs and benefits lead us to think in more nuanced ways about when recording conversations should be permissible, just as many of us have learned to think differently about email privacy at home and at the office? The wide diffusion of writing required standardization to facilitate mutual intelligibility. Will increasingly broad dissemination of spoken language accelerate the demise of regional dialects and less widely spoken languages? Written contracts today have greater legal standing than verbal ones; will that distinction persist in a world in which spoken and written words have equal permanence? How can we harness this new technology to accelerate access to new knowledge, and what would be the implications of the resulting compression of innovation cycles?

Our parents complained that our generation relied on calculators rather than learning arithmetic. Will we complain when our children rely on speech-enabled systems rather than learning to read and write? Near-universal literacy has been one of our greatest accomplishments, with 82% of the planet’s adult population now able to read and write. But it was the ephemeral nature of speech that gave rise to this imperative for literacy, and it is intriguing to imagine what will happen as that imperative abates. Plato said of writing, “If men learn this, it will implant forgetfulness in their souls; they will cease to exercise memory because they rely on that which is written…”. It would naturally have been difficult for Plato to imagine all of the ways in which writing would be used for so much more than as a mere augment for memory—from an Internet that transport ideas through time and space, to great works of literature that transport our imagination to places that don’t even exist. What would a modern-day Plato have to say about the rise of speech to stand alongside writing as a cornerstone for our society? Our generation will unlock the full potential of the spoken word, but it may fall to our children, and to their children, to learn how best to use that gift.

References

G. Bell and J. Gemmell, "A Digital Life," *Scientific American*, 296(3)58-65, 2007; available at http://www.sciam.com/article.cfm?id=a-digital-life&colID=1

J. Hooker (ed.), *Reading the Past: Ancient Writing from Cuneiform to the Alphabet*, British Museum Press, 1996.

P. Lieberman, "The Evolution of Human Speech," *Current Anthropology*, 48(1)39-66, 2007.

W.-H. Lin and A. Hauptmann, "A Wearable Digital Library of Personal Conversations," JCDL, 2002. Available at http://lastlaugh.inf.cs.cmu.edu/alex/JCDL02-RememberingConversationsFinal-r1.pdf

B. Nisbet, N. Yezhkova, L. Connor, "Worldwide Disk Storage Systems 4Q07 Update," IDC 211353, Mar. 2008. Summarized at: http://findarticles.com/p/articles/mi_m0EIN/is_2008_March_6/ai_n24376576

Plato, *Phaedrus* 275a.

J. Rosen, "Researchers Play Tune Recorded Before Edison," New York Times, March 27, 2008. See also: http://www.firstsounds.org/