

Colin Portnuff

AAC – A User's Perspective
October 18, 2006

David Beukelman We wish to welcome Colin Portnuff to the AAC-RERC Webcast series. This is a talk that was presented by Colin on October 18, 2006 at the Oregon Health and Science University. We wish to thank Melanie Fried-Oken and her colleagues for video-recording this presentation so that it can be offered as a webcast. And now, here is Colin, with *AAC: A User's Perspective*.

Colin Portnuff Thank you, Dr. Fried-Oken. Thank you for inviting me to your luncheon meeting. Since this is a luncheon meeting, I trust you will forgive me if I eat as well.

Slide 2 and Pause

Slide 3

I was very pleased to be invited to speak to you today. I spent the most interesting 10 years of my career working closely with scientists and engineers on product development in the medical electronics and software arenas, and I think each of you has the opportunity to contribute greatly to the quality of life of people with disabilities. At Hewlett Packard Company, one of the core company values was that products had to make a contribution to their users, to their field of endeavor, or to the communities in which they are used. I'll come back to this idea of contribution later.

Slide 4

I am going to spend the next 45 minutes or so talking about myself, biases I have noticed, amyotrophic lateral sclerosis, the impact of electronic communication, critical environments for testing, my needs, and your potential to contribute. If you think 45 minutes is a long time to talk, think about this for a second. Every one of the 5,050 words I say to you today was typed by my failing hands. I'm not saying this for sympathy, but rather to give you a quick dose of the reality facing people who use the science that you are building.

Slide 5

I have seven patents for various products that I worked on. All of them were assigned to my employers, so I never earned a cent from them, but I at least have bragging rights to something that you all use everyday.

Slide 6

The page indicator on the scroll box in long documents in windows applications. We invented that at Hewlett Packard company back in the eighties, and implemented it as a time of day indicator in a scrolling 24 hour electrocardiogram display. My other personal claim to fame was that I led the project to develop and sell 35 million dollars worth of defibrillator monitors to the U.S. army, navy, and air force. That was by an order of magnitude the largest contract ever for that type of product anywhere in the world.

Slide 7

I also opened a restaurant downtown, managed it, and sold it on June 1st of this year.

Slide 8

That is the part of the speech called establishing your credentials. That is mission critical for a person with a visible disability, since the first inclination most people have is to generalize from one disability to others. For example, shouting at a blind person, or speaking very slowly to a person in a wheelchair.

Slide 9

These biases sound ridiculous when you lay them out there in the light of day, but believe me, it is a rare occasion when I meet someone and they see me first, not the wheelchair, and a rare occasion indeed when they talk to me, instead of about me. I wish I had a nickel for every time my companion is asked "Can he hear me?" Or in a restaurant, "What would he like?" If I am alone, the general assumption is that I am deaf or retarded, or both. If I have my laptop on my lap, assumptions change. It seems to act as a badge of authority, somehow creating a bridge of normality, at least until I use it to speak. Then the usual reaction is bewilderment. Just when they thought they could relate to me because they use laptops too, I start using it as a speech generating system, and all bets are off.

Slide 10

Does anyone here remember a book called *Black Like Me*? John Howard Griffin wrote about the experience of being a black man in the deep south in 1960. The twist is that he was a white man who dyed his skin dark to learn about racism in America. A couple of years ago, I visited a 96 year old dear friend. She told me, with a brilliant twinkle in her eye, that she woke up every morning, looked around, came to terms with the fact that she had not died in her sleep, and said. Oh, shit. Not again. I sometimes wake up in the morning, and just for a fraction of a second, have to realize that I still have ALS and still can't speak. I feel a bit like John Griffin, only in my case the dye is not going to wear off. And then I think of my old friend, smile, and get on with the day. If my experience is of any value to you, maybe it is because I have a foot in both the speaking and non speaking worlds. That is because I lost my ability to speak only within the last year. In the spring of 2004, I was training for my fourth marathon. I noticed that my speech was getting garbled when I was running. This was a bit of a curiosity to me, nothing more, really. But by the summer, it was quite pronounced, and I began having difficulty swallowing. That combination of speech and swallowing issues got my doctor concerned, and I began to see a neurologist. After exhausting all the other possibilities, he diagnosed my illness as ALS, amyotrophic lateral sclerosis, or Lou Gehrig's disease.

Slide 11

ALS is a degenerative disease in which the motor neurons fail. The upper motor neurons carry impulses from the brain to the spinal cord, and the lower motor neurons carry impulses from the spinal cord to the voluntary muscles. In most cases, the motor neurons affecting limb muscles are the first to go, resulting in weakness, atrophy and paralysis of the arms and legs. Eventually the muscles involving speech, swallowing and respiration follow, and death results from respiratory failure. In a smaller number of cases, about 25%, the muscles involved in speech and swallowing go first, and limb paralysis happens later. This rarer order of progression is known as bulbar onset ALS, and it is the form of ALS that I have. An even smaller number of people will die of respiratory failure before extensive limb involvement occurs. In a small number of cases, usually elderly patients, dementia accompanies ALS, but most ALS patients do not have cognitive deficits. I have always been gifted with good communication skills and a reasonably quick wit, while physical fitness followed only in the last few years, at least as far as my adult life goes. The great irony for me is that when ALS struck, it struck my communication faculties first, and left the physical deterioration for later. I was never a fast runner, but I could and did run 26.2 miles and bench press 200 pounds. Now I can walk about a block and I can't bench press air.

Slide 12

Communication without voice is a difficult challenge. There was a period when e-mail acted as the great leveler. People who could not speak could correspond via e-mail asynchronously. It didn't matter how much of a struggle it was to enter text, because we did it on our own time. Then along came chat capabilities, and we were back to disabled. And then the nuclear bomb hit. Voice chat. Just when we thought it was safe to turn to our PC for a level communication field. Damn. Of course, I'm overstating that. E-mail is still a tremendously useful tool. But speech synthesis is a critical capability not only for face to face communication, but for telephone communication and video conferencing and voice chat, and, well, you get the idea.

Slide 13

There are two critical test beds for speech synthesis products. One is the telephone. How intelligible is the synthesized voice on the phone. On a cell phone. On a less than perfect line? On speaking numbers or spelling a name? The other testbed is a noisy bar or nightclub. Yes, people who can't speak do like to participate in social activities. We don't just need our voices to talk to doctors. It has surprised me to realize how loudly people must talk in social settings to be heard over the background. So before you dare to think your product is usable, live with it as your only way to speak for a month. And accept no limitations on your activity. I had my ALS diagnosis before I lost my speech, and it would have been terrific to bank my voice for the development of a synthesized voice. I noted on your website that there is a group here at OGI working on an engine that will create a voice from a sample of 50 phonemes. What a boon that would be. Every adult should have that sample stored in their voice, against the chance that they might need a synthesized version of their voice at some time. There might even be some neat everyday applications for that voice for people who are not disabled, although I can't think what that might be off hand. A game that parents create for their kids, maybe? A voice response system for the home telephone? And while I'm on the subject of telephones, let me share with you a little more about the difficulties I have on the phone and the strategies I've developed to try to deal with them.

Slide 14 to 17

The telephone is perhaps my greatest source of frustration. Some conversations go very smoothly. My greatest victories are the rare brief conversations where the calling party did not even realize that I was typing to talk. But that is the rare exception. I used to feel the same way about fooling a native French speaker with my perfect accent, which held up only as far as my limited vocabulary took me. But most of my phone conversations are difficult. I am often hung up on by people who just can't figure out what is going on. This happens most often with busy doctors' offices, where triage nurses handle hundreds of calls a day, and if I am too slow to talk after being on hold for ten or fifteen minutes, they hang up on me. I think you can imagine my frustration with that. I am learning strategies that help with phone conversations. Some of my conversation partners like to have my computer click when I am typing, so they know when to be patient and wait for me to say what I'm trying to say. The only drawback of that is that it slows the pace of conversation, because I can't hear them while I am typing, which means I can't type ahead. The phone situation is improving, as I gain familiarity with what works and what doesn't, and also get more comfortable with using quick shortcuts in the system. In outgoing calls, I am working on ways to require the answering party to say something, so they cannot hang up. What I started with was, hello, this is Colin Portnuff. I use a text to speech system to talk, so please be patient while I type. Can you understand me? Ok? But people often assumed that this was a computerized solicitation, and hung up immediately. I have begun to try asking a question first, and then explaining. So a conversation might go like this.

May I have customer service, please?

[female voice] I will transfer you.

[auto call attendant voice] Your call is very important to us. Please stay on the line and your call will be answered in the order received.

[female voice] Customer service, may I help you?

[Colin speaking] Are you the person who can help me with finding out order status?

[female voice] Yes, sir.

[Colin speaking] I use a text to speech system too talk, so please be patient while I type. Can you understand me Ok?

[female voice] Yes, how can I help you?

[back to Colin presenting] So far, this strategy seems to be working better. In a social call, it is simpler. I just start with, Is Mary there? Even if I know it is Mary who answers, she will still have to say, <voice required="name = Crystal16">this is Mary, and I can then identify myself and explain, if she doesn't know how I speak these days.

In both cases, the key is that the answering party has been forced to commit to the conversation at least long enough to figure out that I am not a solicitor or crank caller. One area where I still am struggling is in answering incoming calls. The challenge is that the phone ringing is an urgent interruption, and if I don't happen to be set up and ready, it is impossible to answer the phone and let the calling party know that I have done so. For that I am trying various solutions, including a pocket-sized voice recorder, to tell the caller that I am there and setting up my system too talk to them. So far I have had limited success, but I think I'm gaining ground. I've just changed my message to let the caller know that I can hear them talk while I get set up. So enough about the telephone. What about one to one and one to many conversations?

Slide 18

First of all, one to one conversations. This is the easiest form of communication with this type of system. In a one-to-one conversation we have the great advantage of total commitment to the conversation. There are a minimum of interruptions, and full attention is given. Or at least can be given. The challenges here come when there is strong emotional context to the conversation. First of all, the speech system always has the same, sometimes slightly peculiar intonation. So the listener has to listen for the actual words and ignore the intonation. This is difficult for some people to do. By the same token, there are times when intonation would help greatly to soften the impact of words. I have gotten into hot water a few times saying something that I might have gotten away with by moderating my tone of voice. I am learning to try to use facial expression and gesture to help with communication, and as much as possible to maintain some eye contact and not look at the screen or keyboard while I am typing, although that is difficult. Physical positioning is important. I usually feel uncomfortable when someone sits next to me or behind me and reads my words. It feels like an invasion of privacy, somehow. As if I'm thinking while I'm typing, rather than speaking while I am typing. If I have privacy, I have the opportunity to change what I am saying. This is important, because I sometimes take a bit of a risk and start typing ahead while my partner is still speaking. This often helps speed up the conversation, and if what I am typing turns out to be out of context or inappropriate, I can delete or correct it as they continue and finish speaking. If they read as I type, it takes away my ability to get ahead, slows down the conversation, and at worst, catches me saying something I don't want to say. Another problem with people reading over my shoulder is that they guess ahead, and to paraphrase my friend Michael Williams, who also uses augmentative and alternative communications technology, most people just don't have the horsepower to fill in my vocabulary. The other issue with positioning is that the p c screen itself can form a wall between me and my conversation partner. Melanie has suggested sitting at a 45 degree

angle with my partner, where possible. That works quite well. The other thing I have tried to do is to lower the screen so that it is less of a barrier. One to many communications take two forms, prepared and conversational. This system works quite well for prepared addresses, like this one. I can control the output sentence by sentence, and I can interject comments by typing. Of course, it is tricky to anticipate all the directions you might want to go and prepare for them, but with experience and good area knowledge, it is possible to anticipate some of the things that might happen. Group conversations are one of the great challenges for me. Group conversation is fast and unstructured, and it is very difficult to participate when my pace is so much slower than the group. So I end up typing quite a few things that never get voiced. Occasionally I guess right, and am able to contribute appropriately, but for the most part I don't speak in groups, unless attention naturally devolves on me or I am asked a direct question. That is Ok in social settings, but in meetings it can be a significant handicap. I am fortunate to have the advantage in most meetings of being the one controlling the meeting.

Slide 19

Noisy environments also present a challenge. I have found that in very noisy environments the volume is not high enough to be audible. I have experimented with a LightWriter, which has a display that the listener can read, but my problem with that device is that it is too slow. I now have a small public address speaker that has great volume capability. In some cases I have resorted to turning the display on this system around so that others can read the screen, and that has worked surprisingly well, except for the fact that I can't see my typographical errors, which causes some amusement and confusion. I have some difficulty retaining my sense of humor about typographical errors that result in mispronunciations. I know it is a natural response to laugh at fumbles in speech, but for me they are just frustrating. And speaking of humor, that is another area of difficulty. It is very difficult to get the timing right for humor, or to get the necessary intonation. So I have to be very judicious in the kinds of things that I say for laughs. Word plays and some puns still work, but jokes typically fall flat.

Slide 20

As you can imagine, composition speed is very important. The speed with which I can compose and speak is a critical factor in being able to participate socially and professionally. A great deal of attention has been paid by device and system developers to improving speed, but the state of the art remains unsatisfying for people with limited mobility. Normal human speech occurs at speeds in excess of 150 words per minute, but many people who use AAC are speaking at one or two words per minute. Can you imagine how difficult that would make conversation? Fast typists can compose messages at 40 words per minute or more. But that is a rarity among users of AAC systems. Proponents of morse code interfaces claim that speeds of up to 30 words per minute have been achieved when used with word prediction tools. And one unique system which I will show you a bit later allows about the same speed to be reached.

Slide 21

So aside from the obvious composition speed issue, what is important to me? I see the problem as a Maslov's Hierarchy of Needs kind of thing.

Slide 22

At the base of a pyramid is SAPI 5 compatibility. If I can't use the voice on my system, it is of no use to me.

Slide 23

Next is intelligibility. If I can't be understood, nothing else matters. Once you have achieved intelligibility, you can look to the next layer up in the pyramid.

Slide 24

Next up for me, and closely related to intelligibility, would be pronunciation editing. No matter how good a developer you are, your system will undoubtedly mispronounce some words and many names. Let me set the pronunciation and emphasis for those words. I don't have much of a sense of humor for mispronunciation most of the time. I spoke articulately and with good elocution before last year, and I want that capability now.

Slide 25

Next, I want pitch and speed controls that don't sound totally bizarre. A natural human speaker can vary the pitch and speed of speech without sounding like a machine. But the current pitch and speed controls are almost useless because they cause so much distortion that the voice becomes unintelligible. Why is this important? Well, it's important for me because my wife is hard of hearing, and pitch is an important aspect of intelligibility for her.

Slide 26

Expressiveness would be next. I want a question to sound like a question, and an exclamation to sound like an exclamation. I want to be able to sound sensitive or arrogant, assertive or humble, angry or happy, sarcastic or sincere, matter of fact or suggestive and sexy.

Slide 27

Multilingual capability is next up the pyramid. I used to speak reasonably fluent and perfectly accented French, and some Spanish. I want to be able to speak other languages than English, and in my selected voice.

Slide 28

Loudness. I want a "shout" capability that is not the volume control on my speaker. Sometimes I need to get someone's attention, but once I have it, I want to return to my conversational voice. Setting volume is very difficult. I feel like the Verizon commercial, only instead of, can you hear me now, I'm saying, is this too loud? How about now? How about now? The ability to shout could be life saving for someone with children. Next up is one that I have not heard mentioned, but when I raised the issue on a list serve for users of a.a.c. systems, it seemed to resonate with other users as well.

Slide 29 to 31

I want to talk to animals. Dogs and horses in particular. They do not associate my synthesized voice with me. I don't know if it is a spatial issue or a tonal issue, but they do not respond to the voice at all. I don't know if this is true for all animals, or just for animals that knew me before I lost my voice, but I am afraid it is the former. For me this is a source of sadness, but if you use an animal for assistance, it could be critical.

Slide 32

Finally, at the top of my pyramid would be the ability to sing, in my selected voice, with good timbre and naturally. And without being a musician. I used to sing with perfect pitch, but I could not write sheet music.

Slide 33

The most important piece of advice I can give you is to listen to the voice of the customer in every form that you can find it. But don't be content with short term success from just giving customers what they ask for. My favorite saying from the basic science research organization of Hewlett Packard Company, was, HP Labs, where the rubber meets the sky. It was our gentle way of poking fun at the esoteric research that was conducted there and occasionally bore fruit in the form of usable technologies. Lasting contribution

comes from using your perception to understand what customers need and from using your ingenuity to provide better solutions than we have imagined. Here is an example. There are many systems that employ various approaches to providing alternative access to speech generation for people with limited mobility. There are systems that use on-screen keyboards with a variety of pointing systems, such as mice, joysticks, head mice, and eyegaze systems. There are systems built on iconic representations of language. All solutions that any sensible user could and probably did request. But there is one out of the box approach developed by a group at Cambridge University called Dasher, which offers a completely different way to input text for written or spoken applications. Dasher makes the whole alphabet available with very little motion, can be used with any pointing device, and in the hands of practiced users can generate thirty words per minute. I am looking to Dasher as my post typing modality of choice. 10 Let me take a few minutes to show you a demonstration of Dasher.

Slide 34 and pause

What that demonstration did not show is that Dasher addresses the SAPI and can be used as a direct speech generating interface in a windows based personal computing platform. I have talked about a wide range of issues, from my own perspective as a person with ALS and a user of augmentative and alternative communication systems. Some of them are related to speech engines, such as intelligibility, pitch controls, and expressiveness. Some are user interface issues, like ways to accelerate the speech generation process and alternative input methods. I have not distinguished between engine and interface issues, since as a user that is not important to me, and because I am sure you can readily place them in the appropriate categories. As scientists, you are good at that. But what you may not be as good at is understanding in your heart how terribly important the decisions you make are for hundreds of thousands of people in this country alone. I have spoken from my own experience as a person with ALS, but we are relatively few in number among those who use augmentative and alternative communication systems. There are about 30,000 Americans with ALS. Compare that with about 500,000 people with cerebral palsy. I don't have numbers for people who use AAC who have autism, throat or mouth cancer, stroke or traumatic brain injury. Each person who uses AAC has their own set of requirements. These can vary widely at the user interface level, but at the voice and speech engine level there is a great commonality of need among us. Much of what I have had to say today is related not to speech, but to voice itself. I would ask you to reflect deeply on how we come to associate voice with identity. I have experienced this in a positive way, as people compliment me on my voice. I have heard from several physicians and speech pathologists that my voice suits me. This seemed initially to me to be somewhat preposterous. To me it is not my voice at all, but rather a tool that I employ to allow me to speak. But my family, friends, medical team and acquaintances have integrated the voice as a key part of my identity. In fact, my teenage daughter Lindsay is troubled when I change voices, or even when I correct some of the mispronunciations that she is used to and even has come to enjoy. For instance, good luck Lindsay, instead of good luck. I guess I am beginning to identify with the voice myself, but I would still not hesitate to toss out the voice I use if I could get a more expressive one without sacrificing intelligibility. It is only natural to associate voice with identity, but I think the professionals doing, and guiding, research should be cautious about the flip side. Do you really hear the individuality of each speaker who uses the same voice? As scientists, I know you hear the words and analyze content, but how readily can you see through the artificial characteristics of our voices to the reality of our character and the emotions that we try to express. Can you distinguish clearly between on the one hand, how articulate we are and how much like you we sound, and on the other hand, the actual words and ideas we express? That is, to separate out the quality of the voice from the speech it enables. I would caution you not to make the same mistake I often made in product development. That is, relying too much on consumers I liked, and failing to always,

always continue finding more input, and continuing to ask the same question until I had heard every relevant answer I could find. How easy it was to stop asking at the point I found a few people who validated my own mistaken viewpoint. So while I am gratified by the attention and courtesy you show me, please don't take my views as gospel. The fact that I use Augmentative communications does not mean I am not full of crap. at least on occasion. Until last month, I did not even know this center at OGI existed. I am so very pleased to have met with you today, and I hope there will be more opportunities for us to visit and perhaps work together in the future. Now here are the seven words you've been waiting for. Let me close now with this thought.

Slide 35

I'd like each of you in this room who are engaged in the science of speech and voice development to adopt as your mentor a person or community with impaired speech. While we may not be the mass market for commercialization of your work, if what you do works for us, it should work for any application. Look to the ALS Association of Oregon and Southwest Washington, the MDA Society, United cerebral palsy, or groups associated with traumatic brain injury, stroke or autism. Spend time with us. Learn from us, and teach us. Share what you learn freely and openly with your colleagues. And hopefully, the rubber will occasionally meet the road, and your contributions will have a magnificent impact on someone's life. And when you help someone communicate, you are not just helping that person, but all the people with whom he or she interacts. That is contribution, with a capital sea.

Slide 36

Thank you for your kind attention.