DESIGN OF VOICE CONTROL INTERFACE FOR A MICROWAVE OVEN

Suzanne Keilson
Loyola University, Maryland, USA
skeilson@loyola.edu

Abstract

The microwave oven has become an essential household appliance which can significantly aid independent living for the elderly or persons with disabilities. By its very nature it provides flexible and convenient cooking options without the safety concerns of either gas or electric burners. However, the current user interface design, with flat panel membrane controls and relatively complex sequences of panel contacts for certain commands, can become an almost insurmountable obstacle, especially for those with low vision. A small microphone, Arduino microcontroller, and Sensory, Inc. RSC 4128 chip were all incorporated into an off-the-shelf microwave oven. Arduino outputs were sent to the appropriate membrane panel contacts, based on a limited set of speech commands processed by the Sensory RSC 4128 chip.

Introduction

This paper presents the proof of concept of a voice-controlled microwave oven intended especially for use by persons with visual or motor disabilities. This was part of undergraduate research in reverse engineering and modification design, as well as issues surrounding accessibility of consumer products.

A number of social and market factors are encouraging the development of voice control interfaces for consumer electronics. One of these factors includes the aging of the global population and the desire for people to age independently and in place. The social networks that once allowed people to age in the context and support of extended families are weaker in both developed and developing economies. It is estimated that globally 32.4 million people are blind and 191 million people suffer from moderate to severe visual impairment as of 2010, without regard to age [1]. The elderly are more likely to suffer from depression along with diminishing sensory and motor capabilities. The elderly who are able to maintain their home and social network fair better than those who do not [2]. Often the ability to cook or heat up food for oneself can make all the difference in maintaining that independent life. This spurs the need for appliances and devices with flexible interfaces to accommodate a variety of sensory inputs from and outputs to the user.

The research and development necessary for improving human-machine interactions, which can benefit diverse populations and markets, is often first raised in the context of persons with disabilities. In recent decades legislative norms have established some standards for the development and use of accessible devices. Much of that work has involved two areas. The first is the standardization of web design to make pages understandable to the listener when they pass through a text-to-speech synthesizer [3] [4]. Those requirements for accessible web pages has created significant changes in software for the web [5]. The National Federation of the Blind has been a leader in this area, working with such designers as Raymond Kurzweil. Their development of handheld text-to-speech machines was an instrumental technological keystone in the development of personal digital assistants that extended from the late 1970's until today [6] [7]. The principles of accessibility have also strongly entered website construction and software, including with the force of legal statutes requiring web pages to be readable by those with low vision. Section 508 of the Rehabilitation Act (1973), the Americans with Disabilities Act (1990), and the World Wide Web Consortium's Web Content Accessibility Guidelines (WCAG 2.0, 2008) all direct these requirements [8].

The second is improved physical mobility in the home [9]. Universal design principles were introduced in the layout of living spaces, to address accessibility challenges for persons with mobility impairments. This has resulted in, for example, wider hallways and doorways, lower or adjustable height counters and tables, ramps and other ways to renovate and retrofit multi-level dwellings for easier access. They were developed in 1997 by a group led by Ronald Mace at North Carolina State University [10] [11]. They include seven principles or characteristics of usable products and spaces. The principles are shown in Table One. One of the most famous examples of positive unintended consequences from such universal design is the case study of the curb cut. The curb cut was established for mobility access, but has proven useful to those riding bicycles, pushing strollers or wagons. [12]

Table 1: Principles of Universal Design

Principle 1: **Equitable Use**. The design is useful to people with diverse abilities.
Principle 2: **Flexibility in Use**. The design accommodates a range of preferences and abilities.
Principle 3: **Simple and Intuitive Use**. The design is easy to understand regardless of prior knowledge.
Principle 4: **Perceptible Information**. The design provides feedback effectively to the user in multiple sensory channels.
Principle 5: **Tolerance for Error**. The design minimizes hazards and risks from unintended actions by the user.
Principle 6: **Low Physical Effort**. The design can be used comfortably with a large range of operating forces and minimum fatigue.
Principle 7: **Size and Space for Approach and Use**. Appropriate space is provided for approach, reach, manipulation regardless of user's mobility.

The first principle establishes the mission statement of Universal Design. The second through fifth principle are those of good product design and encapsulate what is needed to bridge the gulfs of execution and evaluation [13]. Principles 6 and 7 apply most readily to the concerns of people with motor and mobility disabilities, but all of the principles focus on the interaction between product (or process, service, space) and the user.

The phrase and the general ideas of Universal Design were transferred beyond the realm of architecture by educators concerned with accessibility and legislation such as the Americans with Disabilities Act (ADA) [14]. The term is used in the education field when referring to lessons plans and learning environments that are aware of and take advantage of the variety of modes of thinking and learning that people display, such as visual, auditory, or tactile learners. The traditional educational system has tended to privilege one set of skills and approaches to learning over others and as a result have disadvantaged some students with learning differences. With the advent of individualized educational plans, things can be tailored to individual needs. Again, as with the curb cut, a very positive, though unintended, result is that many students, who are not necessarily classified as having special needs, also may benefit from a more varied learning environment and approaches. It was hoped that something similar would happen with this project. The primary user we were concerned with was an older person with visual impairment who needs access to a microwave oven to maintain their independence. That does not mean that the product is not of interest to other large market segments, or that with further development this might become a post-market add-on to a consumer appliance or even push a change with some original manufacturers to provide this as a model option.

The first observation that prompted this project was that the introduction of touch-based interfaces, while seen as a boon to some, and an example of an intuitive user interface, actually seemed to be anything but easy and intuitive. It presents a challenge to older users and to the visually impaired. This is especially true for smooth membrane interfaces that provide inadequate tactile feedback. The lack of clear tactile feedback is also coupled with limited visual and auditory feedback and with confusing conventions for accessing the myriad of features made available. As the price point of the microwave oven has fallen and the feature set has increased, limits of adding more features to encourage purchasing has probably been reached and in fact the desirability of a new purchase may even decrease with additional features of marginal utility. This is a design pitfall called “creeping featurism”, identified as a problem by such respected designers as Donald Norman [13]. Rather, looking at ways to modify the sight and motor control necessary for the safe and effective operation of such interfaces represents an opportunity for desirable, feasible product design.

Figure 1 [15] shows a schematic representation of a flat membrane contact incorporated into a printed circuit board. The schematic represents the essence of a three layer approach consisting of a membrane with graphic overlay and conductive contacts, a spacer layer, and the underlying layer with conductive traces. There are options which provide both tactile and auditory feedback via a fourth layer or dome switch. The dome can be made of metal or polyurethane and provides a ‘snap’ as it is depressed and the underlying contact traces are brought into contact, but these are often not chosen for consumer kitchen appliances. Metal domes are more expensive and have a longer life than plastic ones. Both add manufacturing complication and costs.

Figure 2 [16] illustrates the way a flat membrane key gets incorporated into a full, in this case 4x4, keypad. One layer has horizontal traces and one has vertical conductive traces. When the key is pressed, a unique pin on the 8-pin ribbon cable will transmit a signal. Our aim was to provide similar signals via speech recognition on a mixed signal (ADC/DAC) chip and a microcontroller.

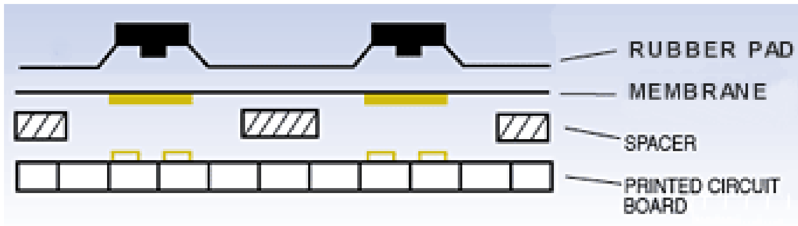

Figure 1. Schematic of membrane key

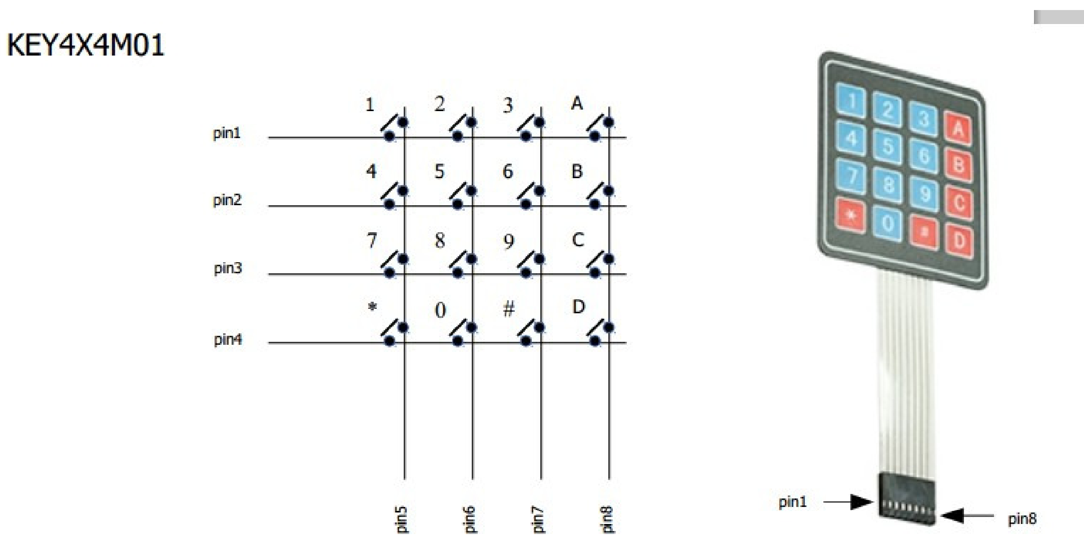

Figure 2. Keypad contact matrix schematic

Design Concepts

The availability of a variety of relatively inexpensive mixed-signal microcontroller chips and software development platforms allows for inexpensive flexibility in design and product development. We utilized a Sensory Inc., RSC 4128 mixed signal processor and VoiceGP module [17].

Some of the significant features of this chip include its use of firmware that incorporates speech synthesis and recognition technologies. The chip works with a proprietary firmware system, FluentChip™, which provides a compact set of speech recognition and synthesis tools. It can record and play three minutes of compressed speech and recognize both speaker dependent and speaker independent vocabularies. Based on an 8-bit microcontroller, the RSC-4128 integrates digital and analog processes optimized for speech recognition. It is intended for exactly this type of small corpus, consumer electronics embedded system application that comes with strong cost constraints. The chip includes a microphone pre-amp, a 16 bit ADC, and a 10 bit DAC on board.

Thus the objective was to build a voice control I/O on top of the existing microwave control to provide proof of concept that such voice I/O can supplement touch interfaces for modest incremental costs for many appliances such as dishwashers, ovens, washers, dryers, and other consumer electronics.

The first branch of design decision is to use speaker dependent speech recognition (SR) or speaker independent SR. Speaker dependent SR requires training of the device, typically with some kind of back-propagation training by feeding back the error signal to the system [18]. In this instance speaker independence is a better option, since the device could easily have multiple users, or even be in a public space.

The next choice was whether to use a restricted vocabulary for the SR or to try a natural language SR approach. The limited vocabulary option was a good fit for the needs of controlling a microwave or most home appliances, where the universe of possible commands is finite and small. Even so, there are a number of challenges in developing a vocabulary for voice control. People will express numbers and time in different ways that have to be accounted for such as "fifteen minutes", "quarter of an hour", or requiring the user to input "one five minutes", which is artificial to a native English speaker. Forcing the user to conform to the machine interface is never the best design option. Another approach in handling this variability in multiple speech utterances mapping to similar meaning and intent is to use ideas from information retrieval. Many samples of actual natural utterances would be incorporated into a look-up index to point to the appropriate control signal [19]. Information retrieval (IR) was originally concerned with problems related to searching a large set of texts and finding those most relevant to a user's query [20] [21]. Two additional problem scenarios are being explored; the case of queries and retrievals from a set of audio data and retrieval from automatic speech transcription. In this case the problem statement would be to categorize and index features of various utterances. The actual utterance would be compared to a set of standard utterances and a measure of its "distance" from each standard calculated. It would then be put into the standard command to which it most closely corresponded and that command's lookup table would instruct the appropriate voltage output.

Even with that potential limitation, it was decided that using the Sensory chip, which can hold a vocabulary of 160 Kbytes or approximately 250 words. The vocabulary to be recognized in the speaker independent mode can be entered as text, which was a huge benefit to this development. The entire memory available was 1.125 Mbytes and contained the control code, the SR words and the list of compressed synthesized machine responses.

In designing the voiced system responses there were other challenges and choices. Feedback to ensure the execution of a command and provide the user with some knowledge of the current machine state is an important principle. However voice interaction is filled with elements of human affect and emotion that is now part of the study of human-machine interaction. Constant machine feedback can be perceived as an irritant; especially to an expert or frequent user of a system. Similarly, an "always on" listening situation for devices may cause confusion in a typical household setting. The general population now can be expected to have experience with automated voice interactions so that more is expected of those interactions and they need to be more fluid and fluent. Rather than having the machine in an "always listening" mode, a large tactile touch button was placed in the side of the microwave. A person with limited vision would be able to feel and press that button and put the machine into listening mode.

Options for voice synthesis include pulling from a limited vocabulary of stored responses from a human speaker or synthesis from an acoustic model. An intermediate approach is to stitch together phonemes and syllables from a library of actual utterances. We chose the first option, to use a recorded library. The design choices made here, while following good practices and usability studies of each design, are things that may set apart each commercial developer for their particular audience and market.

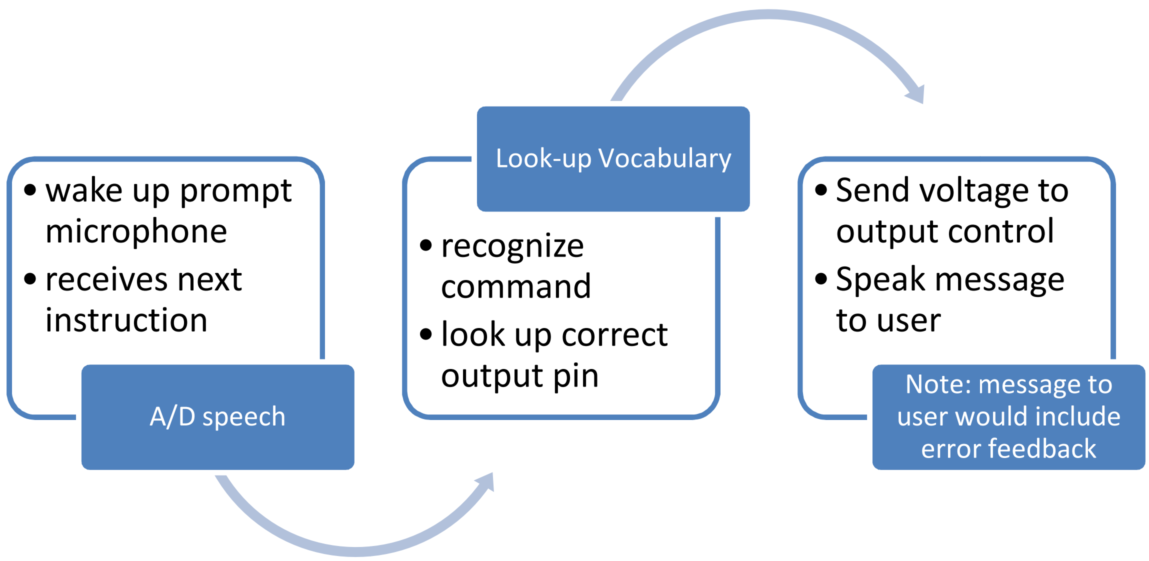

Figure 3. Systems level schematic of voice control.

Prototype

The control signals from the flat panel were mapped to the existing control chip. In the microwave the control signal was a 6x7 dual contact matrix signaling. A subset of the control panel was mapped to optimize the design for the use of the general pinouts from the microcontroller chip. When the user speaks (e.g. "popcorn" or "fifteen minutes") the SR in the chip will match that utterance to the best fit statistical hidden Markov model utterance stored in memory. Figure 2 is a general purpose I/O schematic of the RSC 4128 chip from Sensory, Inc.

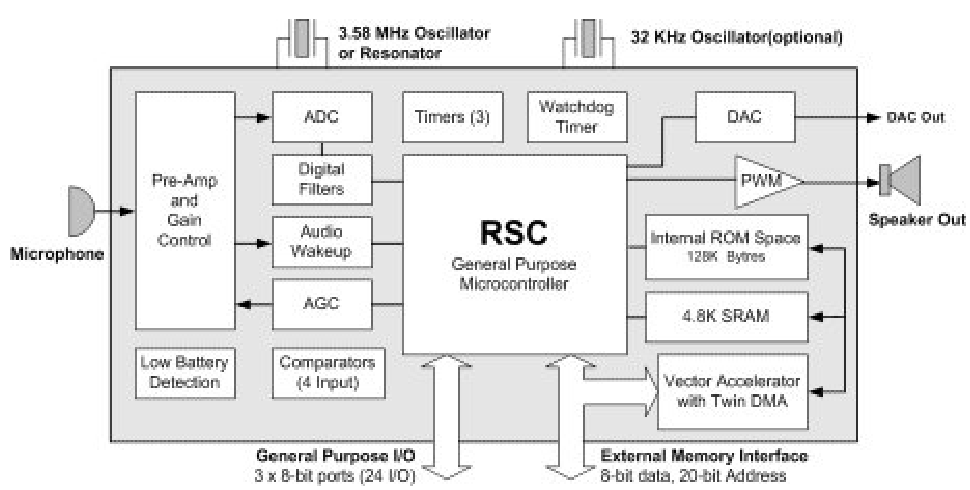

Figure 4. RSC 4128 Block Diagram from Sensory, Inc.

The output signals of the pins of the microcontroller chip are added on top of the controller ribbon and the ribbon's interface with the command and control chip that is already a part of the microwave. This was an important aspect of the reverse engineering elements of the project. If one were to design the prototype from the beginning rather than retrofitting to existing hardware, the preferred design would be to have a single microcontroller integrated on a board capable of both SR and general appliance control. Retrofitting to each model of microwave oven would require the analysis of the control signal mapping for that particular touch pad. The microwave was successfully re-engineered to accept a number of voice commands and execute the correct control signals to the microwave power. This was a successful proof of concept as well as a successful learning experience for the student. The overall costs of this project were modest. The availability of inexpensive, robust mixed-signal microcontrollers makes many such design projects possible, both as individual retrofits for do-it-yourself aficionados as well as potentially commercial viable. Thus this microwave could be controlled in the conventional manner with a touch membrane interface or by (limited) voice commands.

Conclusion

The ability to utilize and integrate robust voice control in consumer electronics and home appliances, with an emphasis on universal design and accessibility, is commercially feasible with currently available mixed-signal microcontroller chips and development environments. Designers and companies need to be trained and encouraged to develop these products. We were able to successfully implement this design, program and test the system to do the following:

- "wake up" with a button press and say "listening"
- Respond to the command "popcorn" by repeating the command
- Respond to the command "popcorn" by starting the microwave
- Respond to the command "x minute" in a similar fashion

Future developments would include conducting usability studies to learn more about preferred modes and styles for the initial group of intended users.

References

[1] Stevens, G.A., White, R.A., Flaxman S. and et. al. (2013). Global Prevalence of Vision Impairment and Blindness: Magnitude and Temporal Trends, 1990–2010, *Ophthalmology,* 120(12), 2377–2384.

[2] Chou, K.L. (2008). Combined Effect of Vision and Hearing Impairment on Depression in Older Adults: Evidence from the English Longitudinal Study of Ageing, *Journal of Affective Disorders,* 106(1–2), 191–196.

[3] Milne, S., Dickinson, A., Carmichael, A., and Sloan , D. (2005). Are Guidelines Enough? An Introduction to Designing Web Sites Accessible to Older People, *IBM Systems Journal,* 44(3) 557.

[4] Mankoff, J., Fait, H. and Tran T. (2005). Is Your Web Page Accessible?: A Comparative Study of Methods for Assessing Web Page Accessibility for the Blind, *SIGCHI Conference on Human Factors in Computing Systems*, Portland, Oregon.

[5] Schneiderman, B. (2009). *Designing the User Interface: Strategies for Effective Human Computer Interaction*. (5th edition), New York: Addison Wesley.

[6] Massof, R.W. (2003). Auditory Assistive Devices for the Blind, *Proceedings of the 9th International Conference on Auditory Display (ICAD)*, Boston, MA.

[7] Kurzweil, R.C., Albrecht, P., Gashel J., and Gibson, L. Mode Processing in Portable Reading Machine. US Patent US7505056 B2, 17 March 2009.

[8] You, C.X., Nicolini, A.M., Ippolito, V.J., Fink, W., and Michalska, A.M. (2014). Accessible Web Page Design for the Visually Impaired: A Case Study, *International Journal of Computer-Human Interaction,* 30(12), 995-1002.

[9] Crews D.E., and Zavotka,S. (2006). Aging, Disability, and Frailty: Implications for Universal Design, *Journal of Physiological Anthropology,* 25(1), 113-118.

[10] Mace, R.L. (1998). Universal Design in Housing, *Assistive Technology: The Official Journal of RESNA,* 10(1), 21-28.

[11] Story, M.F., Mueller, J.L., and Mace, R.L. (1998). The Universal Design File: Designing for People of All Ages and Abilities, Raleigh, NC: Center for Universal Design, NC State University.

[12] Peterson, J. Smashing Barriers to Access: Disability Activism and Curb Cuts, Smithsonian Institution, 15 July 2015. [Online]. Available: http://americanhistory.si.edu/blog/smashing-barriers-access-disability-activism-and-curb-cuts. [Accessed 15 July 2016].

[13] Norman, D. (2014). The Design of Everyday Things, revised and expanded, New York: Basic Books.

[14] Rose, D., Meyer, R. (2000). Universal Design for Learning, Journal of Special Education Technology, 15(1), 67-70.

[15] Epectec.com, Optimal Membrane Switch Design, epectec.com, [Online]. Available: http://www.epectec.com/articles/optimal-membrane-switch-design.html. [Accessed 20 July 2016].

[16] Electronics Stackexchange, [Online]. Available: http://electronics.stackexchange.com/questions/96309/4x4-keypad-to-microcontroller. [Accessed 20 July 2016].

[17] About Sensory: Sensory Inc., Speech Recognition Company, 10 June 2010. [Online]. Available: http://www.sensory.com/products/integrated-circuits/rsc-4x-series/. [Accessed 10 May 2016].

[18] Huang, X., and Lee, K.F. (1993). On Speaker-independent, Speaker-dependent, and Speaker-adaptive Speech Recognition, IEEE Transactions on Speech and Audio Processing, 1(2), 150-157.

[19] Liddy, E.D., Paik, W., M. E. McKenna, M.E., and L. Ming. Natural Language Information Retrieval System and Method. US Patent US5963940 A, 14 August 1996.

[20] Foote, J. (1999). An Overview of Audio Information Retrieval, Multimedia Systems, 7(1), 2-10.

[21] Brown, M.G., Foote, J.T., Jones G.J.F., Jones, K.S., Young, S.J. (1997). Open-vocabulary Speech Indexing for Voice and Video Mail Retrieval, *Proceedings of the fourth ACM international conference on Multimedia (MULTIMEDIA '96)*, New York, New York. 307-316.

Biographies

SUZANNE KEILSON is currently an assistant professor in the department of Engineering at Loyola University Maryland. She has served as Associate Dean of the College of Arts and Sciences at Loyola and in volunteer leadership positions at the American Society of Engineering Education (ASEE). She earned her B.A. in physics from Yale University, as well as an M.S. and Ph.D. in Applied Physics from Columbia University. Her research background and interests include, materials science, signal processing in the auditory system, engineering design, design and engineering innovation and education. Dr. Keilson may be reached at Skeilson@loyola.edu

Acknowledgments

This work was conducted as part of the Hauber Summer Research Fellowship Program by Jamel Garrett, Loyola Class of 2010 and with the invaluable help and technique expertise of Jason McMahon. The Hauber program's endowment funds students across the Natural and Applied Sciences at Loyola University Maryland.