EXPANDING THE HUMAN BANDWIDTH THROUGH SUBVOCALIZATION AND OTHER METHODS

An Interactive Qualifying Project Report submitted to the faculty of

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

____

Christopher M. Ceccarini

January 8, 2008

Project Advisor: Professor Fredrick Bianchi
Project: FB-002

1 Abstract

This is a look at human bandwidth and how it applies to human-human interaction and human-computer interaction. The paper discusses what human bandwidth means and what must be done to try expanding it. Current methods of expanding bandwidth are discussed. The methods include detection of subvocal activity, facial expression detection, eye tracking, emotion detection in digital music, pen based musical input systems, and augmented reality. After explaining these methods, the paper focuses on using some of the technologies together to give an idea of what the future of interaction with computers might look like. These proposed ideas include emotion based music, various uses for augmented reality, and composing music with the mind.

Table of Contents

Table of Figures

Table of Tables

2 Introduction

People encounter many things in a day. Human bandwidth refers to the number of ways in which we can interact with the world. For human beings, this bandwidth is very wide. For example, interactions with other human beings use the most of this bandwidth, taking advantage of things like the voice, hand expressions, conscious facial expressions, and even through writing, typing, or playing music. We can even say there are even more ways for humans to interact with each other that involve things we do not even consider. These methods are handled by the subconscious, also referred to as our adaptive unconscious. This handles things like facial expressions that we are not aware of making, releasing hormones, and different things that may affect decision-making abilities.

This paper will focus on the limited bandwidth of humans as it applies to interactions with other human beings and with objects like computers and musical instruments. One specific breakthrough technology is the detection of subvocal signals. Subvocalization will be introduced, explaining exactly what it is, how it can be detected, and how it is being used today.

The paper will also discuss methods already in use today to expand human bandwidth. One new technology is a pen based input system. This system allows composers to write their music easily into a computer. Newer systems allow for easier manipulation and easier, more compatible input gestures to capture exactly what it is that the composer has in his head. Eye tracking is another field of study and may provide an alternate method for people to interact with the computer. Detection of facial expressions is crucial in determining a person's mood. Music invokes many moods, and the detection of a specific mood is becoming more of a possibility rather than science fiction. Finally, virtual reality has been an idea for many years now. Virtual reality puts the user into a different world, allowing them to look around and manipulate their environment.

The goal of this paper is to use the previously mentioned methods of expanding bandwidth. It will also look at the direction in which interacting with computers, playing musical instruments, digitizing music, and even just listening to music is going. Imagine a world where all you would have to do is put a device on your head and play a computer game simply by mimicking the gestures you would want your character to make. Maybe just by sitting in front of your computer, your computer would detect what mood you are in, and play a specific set of songs to brighten your day. For composers, the ideal way to get a song recorded in digital form would be to simply think of the song in his/her head and have it transposed to a digital file on their computer.

3 Limited Bandwidth

Each person has countless methods of interaction. In other words, bandwidth is very wide and unlimited. As previously mentioned, humans have a voice, hands, face, and even other body parts that can be used in communication. This section will discuss bandwidth and how it is limited in respect to interaction with other people and interaction with computers.

3.1 Person-to-Person Interaction

When interacting with another human, there are different ways to convey a message. This is done by using the voice, facial expressions, even particular hand and body expressions. It is a well known fact that people communicate with each other even when no words are spoken. The following is a quote by Sigmund Freud dated 1905:

> *He that has eyes to see and ears to hear may convince himself that no mortal can keep a secret. If his lips are silent, he chatters with his fingertips; Betrayal oozes out of him at every pore. (Freud, 1963)*

When combined, all behaviors paired with all of the senses provide humans with a very powerful, high bandwidth method of communication.

3.1.1 Voice

When considering voice, that in itself has a particular bandwidth. First, when talking to somebody, the amplitude of speech can be changed from either a whisper to a yell. Tone of voice can also be communicated to differentiate between emotions like anger, compassion, fear, sadness, happiness, confusion, seriousness, and countless other emotions. A person, telling the voice to sound harsher, talk softer, or yell, does some of this consciously. In other cases, the adaptive unconscious does this instead. This includes putting fear into the voice or using a certain tone of voice when talking to females and a different one when talking to males. However miniscule this difference may be, the subconscious is actually a very powerful thing.

The voice is limited though. The human voice is limited to the range of 80-1100 Hz, disregarding anomalies such as opera singers and world record holders. This range of frequencies is well within the range of hearing (20 to 20,000 Hz). The frequency of speech is used for many things, including inflection at the end of a sentence. Would more emotions be possible to express if the vocal range expanded beyond 1100 Hz? Would it incite more fear into people by yelling at them at a frequency of 40 Hz?

Speech is also limited by language in many ways. Somebody who speaks only English can go to China and speak to people, but many of them would not understand at all. It works the same the other way as well. That person would hear people talking in Chinese, but would not be able to interpret it. Granted, bandwidth through speech is very wide, but this does no good if the person at the other end cannot interpret it. Imagine speaking to somebody who only speaks and understands Mandarin. Certain feelings can be expressed through tone of voice and facial expressions. Different things can be explained through hand motions and body movements. However, the actual information is lost due to limited bandwidth.

3.1.2 Face

One thing that every human being pays attention to on another human is their facial expressions. Many things such as a smile, frown, a tongue poke, and an angry face can be made whenever we want by making the conscious decision to contract or relax certain muscles in our face to conjure up the desired facial expression. Somebody can be told to smile or make a funny face. One may also open their mouths when something shocking happens. Facial expressions have no “language barriers”. A smile means the same thing in America as it does in isolated tribes in Africa. Many facial expressions, just like the voice, are controlled by the subconscious. If you were to watch yourself while talking to somebody, frame for frame, you might not believe the number of facial expressions that you make while listening and talking.

The face is limited in that it can only express emotions. You cannot ask somebody to go to dinner just by using facial expressions.

3.2 Person-Computer Interaction

Back when computers just began to take off, the methods of interacting with them was very limited. There was a keyboard, a monitor, and maybe a mouse. As time went by, things like a mouse, scanner, printer, and touch screens became commonplace. Nowadays, many computers come with speech recognition. When games are played, there are joysticks, game pads, and steering wheels. Some computers even use multiple monitors for 3-dimensional visual effects and Dolby® Surround Sound for 3-dimensional audio effects.

3.3 Music

When referring to human bandwidth in respect to music, the aspect that we focus on is the limited number of ways a human can interact with a musical instrument or device. Whether through voice, fingers, feet, lips, palms, or even just knowledge of a MIDI device and how to make music with it, the human body still has a limited bandwidth. Limited bandwidth refers to our ability to think of the way a song should sound versus the ability to perfectly recreate that specific sound. If you think of the human body as a device of various outputs and instruments as devices of various inputs, the problem becomes quite clear: there are not many ways for a human to interact with an instrument. In order to address the issue of limited human bandwidth, it is necessary to look at the current methods of receiving data. Also, advances in technology must be researched and improved in order to expand bandwidth. The ultimate expansion of bandwidth in terms of music would be to simply think of the way a song should sound, and have it instantly created in a digital format. However, with mind reading still far out of reach, more realistic means of capturing the feel and emotion of a song must suffice.

3.3.1 Instruments

When considering the human-instrument interaction, countless examples come to mind. For instance, a guitarist interacts with his output device (usually an amplifier with a speaker or through the resonating chamber in an acoustic guitar) in many different ways. Five fingers on the left hand and one pick in the right hand (for right handed people) act as the primary interaction between human and guitar. The various combinations of different scales and bends with the left hand and different rhythms with the right hand already make for complicated music playing. However, there are other processes before the sound actually reaches the speaker. An amplifier may be changing the volume and can even over-amplify (commonly referred to as overdrive or distortion) a sound. Also, the feet could be used as an input device on different effect pedals such as a wah pedal, chorus pedal, reverb pedal, or even as simple as a volume pedal. Some artists, such as Peter Frampton and Joe Walsh, have mastered the art of using the mouth along with a device called a talk-box. A talk-box utilizes a plastic hose placed in the mouth to detect the shape of the mouth. The result is a sound that makes a guitar sound like it is talking. Other guitarists' attempts to expand bandwidth include using a violin bow on guitar strings (Jimmy Page) or even just pounding the body of their solid body Fender Stratocaster (Pete Townshend) in order to create a truly unique.

Figure 1 - Jimmy Page with violin bow (©Bob Gruen/www.bobgruen.com)

If applied to a piano, a human has a limited output of ten fingers and two feet. The left hand usually plays chords on the lower notes while the right hand, positioned on the higher note keys to play a melody. The feet can be used to add sustain, or in the case of an organ, can be used to add a bass-line.

3.3.2 Digital Music

An aspect of music that many people might not even consider is the time and effort put into digital music. When looking at the differences between having a piece of music printed in a book and one in a digital MIDI format, the scales tip in favor of the digital world. Rather than just being able to play harmonic music that is in time, a MIDI composition can be changed in many ways just by changing values in the computer. Sustain, tremolo, pitch, tempo, and even the instrument can be changed when interacting with a MIDI recording. Advanced music programs can even automatically change the key a music piece is in. Notes can be added, deleted, shifted, and modulated, adding an incredible amount of versatility when recording and editing music.

When an artist wishes to store a song in digital format, there are many options. The most basic method of input is through direct keyboard/mouse interaction with whatever musical composition software, such as Finale, the user decides to use. Another option is a MIDI keyboard, which works by reading keystrokes and placing them onto a digital musical staff. A more recent method involves putting music into digital form by optically scanning a musical composition into a computer program, which will automatically place notes in their proper places in a digital file. The down side to all of this is that it is not portable, can be too slow, and may be difficult for the computer to interpret. What if a user wants to write music directly into a computer? The preferred method of writing music is by handwriting it on a piece of paper, which would make use of the optical scanning method.

4 Current Research on Bandwidth Expansion

When thinking about expanding bandwidth, one must consider the things that make interaction possible. Everything that somebody does to interact with another person or object is able to be captured and measured. When speaking, another person is able to listen and they put their brain in charge of interpreting the frequencies and sounds made. When using a computer, the processor of a computer can capture and interpret the movements made on a mouse or the keystrokes made on a keyboard. In short, in order to expand bandwidth, we need to find other signals that can be measured and interpreted. Technology is expanding very quickly, but is not quite advanced enough to read minds. However new technology in the near future can help expand the ways humans express themselves and how those expressions might be detected and used.

Some possibilities may include different sensors placed on the body to detect heartbeat. Using the pace of the human heart, maybe a controller could change the volume, vibrato, distortion, or some other effect to the original sound.

One "futuristic" instrument that had changed the human-instrument interaction methods was the Theremin. The Theremin used a rod that senses where an object is height-wise and how far away from the rod it is. Depending on the position along the rod, the pitch would raise or lower. Depending on how far away the object is, the volume is either increased or decreased. Most people who play the Theremin use their hands using conductor-like motions to create truly unique music.

This section will discuss some methods of expanding bandwidth already being used. One major breakthrough is the detection of subvocal signals. Subvocalization will first be introduced and then the methods of detecting and using it will be discussed. Along with subvocalization, things like pen-based input systems, eye-tracking, facial expression detection, and virtual reality will be introduced. This chapter will go over the current state these technologies are in and how they are being used today will be discussed. The goal of finding new ways to expand human bandwidth using these will be discussed in Chapter 5.

4.1 Subvocalization

One new technology being researched by NASA is the detection of subvocal signals. Subvocalization is a phenomenon that occurs within everybody. It refers to the actions the brain, tongue, and vocal chords take when preparing to speak. Every time a word is read silently, the brain still prepares the tongue and vocal chords to speak those words. This results in small signals being sent to slightly move the appropriate body parts to speak, but nothing is heard.

Research on subvocalization has uncovered many ways to detect and use this "silent speech".

As previously mentioned, subvocalization is basically silent speech and it happens within everybody. Wikipedia defines the word:

> ***Subvocalization****, or silent* speech*, is defined as the internal speech made when* reading *a* word*, thus allowing the reader to imagine the* sound *of the word as it is read.*[1]

Imagine reading a book (or even just reading this report), every time you look over a word and read it to yourself, you are subvocalizing that word. Your brain is preparing

[1] (Carver, 1990)

your vocal chords and your tongue to speak the words you have just read. All it would take to hear those words is you brain to tell your lungs to release air and pass it over the voice box (larynx).

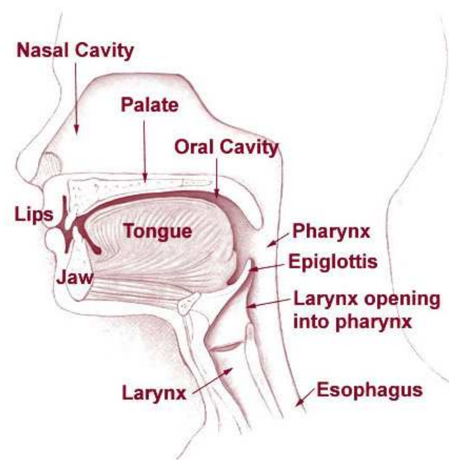


Figure 2 - Diagram of body parts used in speech[2]

After air passes through the vocal chords of the larynx, that air is amplified in the pharynx and further distorted (what we hear as talking) by the tongue. When reading words, the brain prepares for these different actions, but never actually uses air to make a sound.

As readers and speakers of a phonetic language, it is easiest to read words as they are spoken. By subvocalizing, the mind reduces the cognitive load it takes to read words. It is said that it is possible to train a mind to not subvocalize when reading. The method used is to take entire sections and read them as a whole part. For example:

The cat ran away.

[2] (Wikipedia, 2007)

If you were to read the previous sentence to yourself, you will notice that in your mind, each word is spoken separately, and it was not until the end that the entire sentence became clear. In order to train a mind to not subvocalize, that sentence should just be seen together and understood immediately. In other words, a group of words should be seen and read together and the meaning should be understood, rather than reading individual words and using cognitive skills to piece it all together as a phrase. The reason for this training is said to have positive effects on the speed of reading, however, this point is often disputed.[3]

4.2 Detecting Subvocal Signals

The first issue to address when thinking about how to detect a subvocal signal is to determine what kind of signal we are looking for. When somebody subvocalizes, the tongue and vocal chords react so slightly that the best way to detect it is through electrical signals. A subvocal signal is an electrical potential created by the muscles in the larynx and the tongue. The signals are recorded by an Electromyograph (EMG). It is these signals that NASA uses to “read minds”.

When at rest, the signal received from the larynx, tongue, and jaw is recorded to be about 3µV. When a subject speaks out loud, this signal can be about 1mV.[4]

The following paragraphs will describe the methods tried in this project in attempts to at least see these signals.

[3] (Carver, 1990)
[4] (Hardyck, Petrinovich, & Ellsworth, 1966)

4.2.1 BrainWave Device

The first attempt in trying to read these signals was through a device provided by Professor Bianchi. The BrainWave device is designed to connect sensors to it and relay information to a computer for analysis. A few issues arose in the attempts to use it however. First of all, an updated license for use of the provided software was not available for our budget. However, a trial version of the software was available with limited use. The software had plenty of demonstrations along with example applications and placement of sensors on the head of a test subject. Using the BrainWave device on myself produced no readings on the computer. This could have been due to anything: a poor connection with the computer, poor placement of sensors, a faulty device that just does not work anymore, or a brain-dead test subject. The fault was narrowed down to either poor sensor placement or the BrainWave device was not communicating properly with the computer.

4.2.2 Instrumentation Amplifier

The next attempt in trying to view the EMG signals was through an instrumentation amplifier. Being an Electrical and Computer Engineering major, I decided to take a more direct method. How about building a circuit that is used in most hospitals right here in our own ECE department? After speaking with a couple of professors, Professor Bitar pointed me in the direction I wanted to go. He provided me with the following circuit from an ECE2011 lab for an instrumentation amplifier.

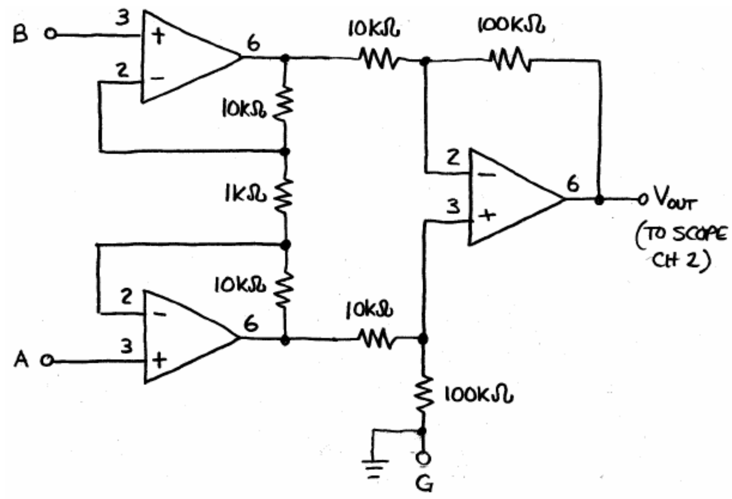


Figure 3 - Instrumentation Amplifier[5]

This circuit uses three LM741 Operational Amplifiers along with a network of resistors to create a proper amplification of the EMG signals. Points A and B were attached to my body and V_{OUT} connected to an oscilloscope. Proper functionality of the circuit was confirmed, so I proceeded to attempt to read a signal from either the tongue or the larynx. Neither attempt yielded any results. The problem was suspected to be a result of not having enough amplification between the input and the output.

4.2.3 Proven Method of Detection

As assumed, the issue with detecting subvocal signals was the lack of proper equipment. Dr. Chuck Jorgensen has been working with subvocal signals for uses with many different applications ranging from the physically or mentally handicapped to NASA astronauts and HAZMAT workers.

Two pairs of AG/AG-Cl electrodes are placed on the throat. Two are placed on the left and right anterior part of the throat about 0.25cm back from the chin cleft. The other two

[5] (Bitar)

are placed about 1.5cm from the left and right side of the larynx. The placement of the sensors is shown in Figure 4.

The sensors were connected to a Neuroscan signal recorder and sampled at 2000Hz. SCAN4 Neuroscan software was used to remove anomalies such as coughs and swallows while MATLAB scripts were used to detect more complex features in the recorded signal. Using software to apply various mathematical transforms, data could be acquired and then used to train a system, a similar technique to the system used by current voice recognition software programs.[6]

Figure 4 - Sensor Placement with Keyboard Example[7]

Using this method of acquisition and training, people were asked to read a set of words to themselves and the results are shown in Table 1.

[6] (Jorgensen, Lee, & Agabon, Sub Auditory Speech Recognition Based on EMG Signals, 2003)
[7] (Hart)

Table 1 – Accuracy of Sub Vocal Detection[8]

	Transform Used			
Word Spoken	Dual Tree Wavelet	Fourier	Hartley	Moving average
“Stop”	84%	83%	79%	62%
“Go”	100%	100%	97%	90%
“Left”	91%	91%	91%	84%
“Right”	80%	89%	91%	91%
“Alpha”	97%	82%	79%	73%
“Omega”	97%	98%	100%	95%
Average	92%	91%	90%	83%

4.3 Using Subvocalization

The focus of this project is how to expand human bandwidth, and as mentioned before, it is necessary to find different, more breakthrough methods of detecting signals from people. Subvocalization is on the edge of becoming something huge, and is therefore a large point of interest for this project. Its uses are already being researched, so the focus turns to how one might use it in regards to music, and how it could be used in the future. Advances in technology within NASA have made extreme breakthroughs in using subvocal “thoughts”.[9] Subvocal activity might also be used to solve Professor Bianchi’s current issue concerning human interaction and limited bandwidth. Future advances in detecting subvocal activity might lead to uses in the armed forces or in homes of elders.

4.3.1 Current Applications

Currently, engineers at NASA, in particular a scientist named Chuck Jorgensen, are researching subvocalization. So far, NASA has in development a way to interpret simple words and digits. By what is reported, they can interpret the numbers 0-9, and also simple

[8] (Jorgensen, Lee, & Agabon, Sub Auditory Speech Recognition Based on EMG Signals, 2003)
[9] (Jorgensen & Binstead, Web browser control using EMG based sub vocal speech recognition, 2005)

words such as “stop”, “go”, “left”, “right”, “alpha”, and “omega”. Using this truncated vocabulary, and setting up a matrix similar to that in Table 2, people were actually able to navigate through web pages just by thinking.

Table 2 - Matrix Example

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	0	1	2	3	4	5	6	7	8	9
1	!	@	#	$	%	^	&	*	(	)
2	Q	W	E	R	T	Y	U	I	O	P
3	A	S	D	F	G	H	J	K	L	:
4	Z	X	C	V	B	N	M	<	>	?
5	q	w	e	r	t	y	u	i	o	p
6	a	s	d	f	g	h	j	k	l	;
7	z	x	c	v	b	n	m	,	.	/
8	Space	Alt	Ctrl	Bksp	Enter	Shift	Tab	C.Lock	N.Lock	S.Lock
9	Del	Ins	Home	End	PgUp	PgDn	ESC	+	-	=

4.3.2 Addressing Limited Bandwidth Using Subvocalization

As previously mentioned, subvocalizing is something everybody already does without even knowing it. Incorporating subvocal signals into music could be used to control many aspects of the way music sounds. A talkbox is a device where it incorporates the shape of a guitarist’s mouth and the pressure applied to a tube and uses that as an input to change the phase, pitch, and overall dynamics of a sound. In terms of a MIDI keyboard, everything about a note or an effect is controlled by a specific MIDI value.

It may be possible to use subvocal signals to act in such a way that it controls a value or an effect to change the way certain notes sound. For example, if used as an input to a MIDI device, subvocal signals might be able to be used as a way of keeping tempo, adjusting the brightness of a note, or even to add vibrato to a note.

A simple concept for a MIDI application, as mentioned earlier, might be to adjust a MIDI control value. Even if the sensors used to detect subvocalization are simply qualitative rather than quantitative, this information could still be used.

4.3.3 Future Applications of Subvocalization

Imagine you are sitting down in a very important business meeting. Your phone rings and you answer the phone. However, instead of saying anything, you are able to speak without uttering a single word. Maybe you are a soldier in a covert operation and silence is critical. Perhaps giving completely silent orders will give our troops the upper hand over our enemies.

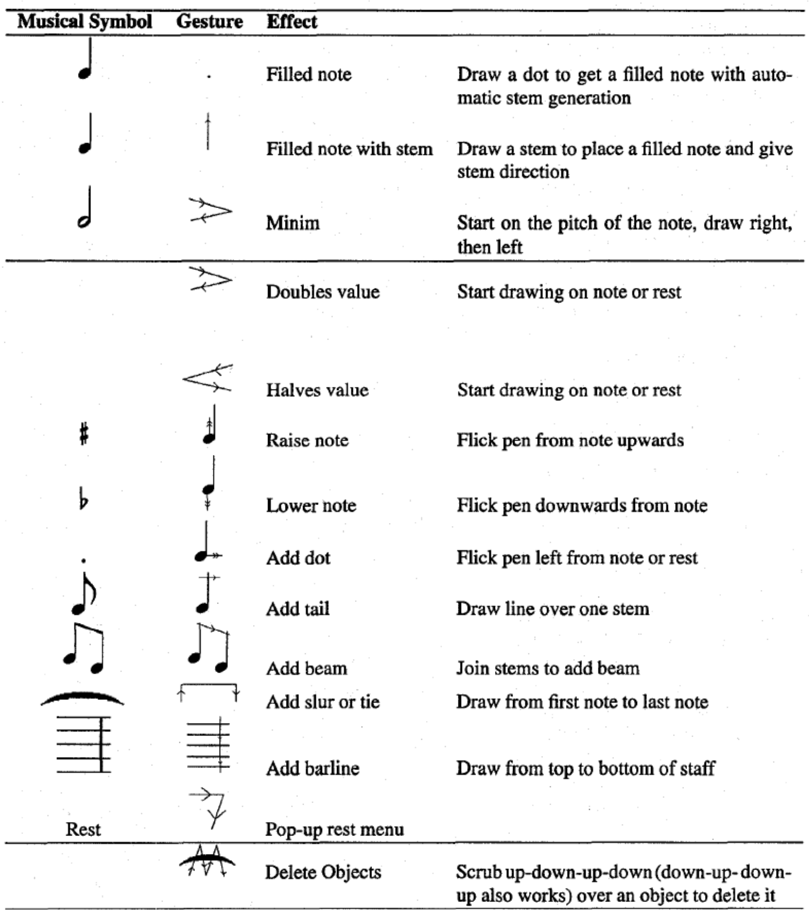
Currently, research at WPI by Professor Rick Brown is using a device to detect the shape of the throat and position of the tongue during speech in order to filter out background noise during cell phone conversations.

4.4 Pen Based Input System

As mentioned before, there are only so many ways a user can create a digital musical composition. A new method being researched is a pen-based musical input system.[10] By creating a set of gestures, as shown in Figure 5, to be used with a program called *Presto*, users can input music directly into a computer as fast as they are able to write it. The whole issue with inputting music directly into a computer is pretty simply explained, but creating an environment in which this is possible may be a little tougher. Current systems are able to convert written words and sentences into a digital word processor. The difference between this and music is that words are essentially two-dimensional. A word has defined letters in it, there will only be one letter at a time, and a sentence has no structure of time or rhythm. With music, there is a completely new set of symbols (half-notes, whole notes, rests, tie-lines, and bars). Music also may have more than one note per position in time. In order to convert written music into digital music would need to deal with issues. Presto has created such a system in that rather than just writing symbols, there are simple gestures to tell the computer which notes are to be written, erased, tied together, and grouped (to make chords).

By using these gestures, users ranging from experts writing from memory and beginners copying music from paper were able to input music into a digital format three times as fast as with the previous methods. Presto is just one way people are looking at to improve the methods of human-computer interaction.

[10] (Anstice, Bell, Cockburn, & Setchell, 1996)



Musical Symbol	Gesture	Effect	
		Filled note	Draw a dot to get a filled note with automatic stem generation
		Filled note with stem	Draw a stem to place a filled note and give stem direction
		Minim	Start on the pitch of the note, draw right, then left
		Doubles value	Start drawing on note or rest
		Halves value	Start drawing on note or rest
		Raise note	Flick pen from note upwards
		Lower note	Flick pen downwards from note
		Add dot	Flick pen left from note or rest
		Add tail	Draw line over one stem
		Add beam	Join stems to add beam
		Add slur or tie	Draw from first note to last note
		Add barline	Draw from top to bottom of staff
Rest		Pop-up rest menu	
		Delete Objects	Scrub up-down-up-down (down-up-down-up also works) over an object to delete it

Figure 5 - Main gestures in the *Presto* system

4.5 Facial Expressions

When a guitar legend like B.B. King picks up a guitar and plays it, he is not simply taking music from a sheet, processing it, and playing it. He is adding a feeling to the music. His guitar playing alone could tell a story. By bending the strings and using natural harmonics, he can create the effect that his guitar is crying. By adding distortion and using lower notes along with the tempo and rhythm kept with the right hand, he can make the guitar sounds angry. By removing the distortion and playing arpeggios with higher notes, a guitar can sound very happy. All of these emotions can be expressed by the human face, and with enough practice, these emotions can be expressed through music.

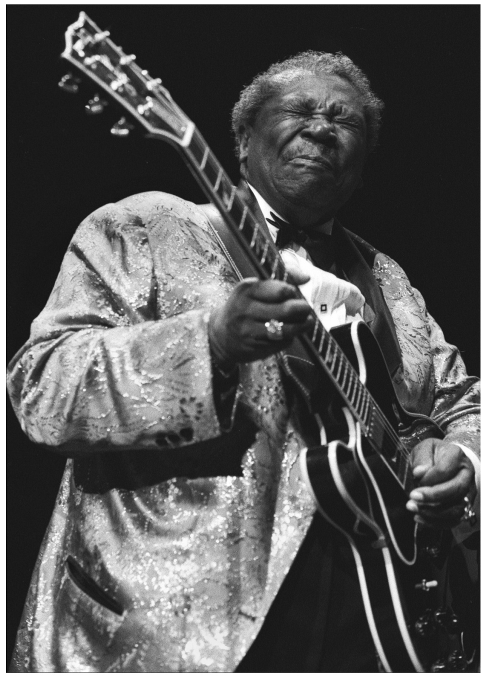


Figure 6 - B.B. King[11]

As seen in Figure 6, B.B. King has a face of pain. When playing physical "analog" instruments rather than creating a digital piece of work, it is much easier to express this type of emotion. Maybe by taking readings of a face while a person is using a MIDI keyboard to input music into a computer or synthesizer, the computer would automatically increase distortion or increase vibrato. By making a painful face, distortion and vibrato would be added. By making a happy face, treble would be increased and distortion removed. By shutting the eyes really hard, perhaps a bend would be added to a note.

Facial features can also be used in our digital gaming world. With things like Massively Multiplayer Online (MMO) games, Xbox 360, and online chat rooms, facial features can

[11] (Brewvana)

be used in a very neat way. An example of an MMO game is World of Warcraft created by Blizzard. Every person in this digital world takes a character and explores while engaging in battles to gain experience. The great part about this digital world is that everybody who is signed on and playing can interact with everybody else in the world. Now we consider what is possible with facial recognition software. If your real time facial expressions could be transferred to your online character, maybe less would need to be said if a person walks up smiling and waving. Or maybe battles could be started if a person frowns at somebody and the other person frowns back.

America Online has one of the most popular means of communication besides the phone and email. An instant messenger client allows people all over the world to "talk" to each other in real time by typing messages to each other. There are many other symbols people use such as:

:) - Happy

:(- Sad

:-P - Tongue poke

LOL - Laughing Out Loud

IDK - I don't know

Facial expression recognition is being developed currently and newer ways are being thought of to enhance this technology. The study of facial expressions has been a huge topic, determining whether facial expressions are universal or restricted to certain cultures. At least six universal facial expressions were found to be universal. These expressions include *happiness, anger, disgust, sadness, fear,* and *surprise*.[12] People from

[12] (Ekman & Friesen, 1971)

cultures all over the world were able to look at pictures of people expressing these emotions and correctly guess the given expression.

One of the newest methods of facial expression recognition is through a cloud basis function (CBF).[13] Some of the problems with facial recognition by computers are that everybody's face is different. Age, gender, and culture play a big role in what a person's face looks like. The CBF works to enhance current facial recognition by applying newer, advanced formulas that automatically take into account the variations among gender, race, and age.

Facial expression recognition would be a great thing to use on a personal computer with music, games, and applications. However, this technology could also be used by authorities to determine whether a person is telling the truth or by surveyors to determine whether a certain movie instilled the correct emotion among people in the audience.

4.6 Emotion Recognition

Emotion recognition in music is a huge possibility. In 2004, it was suggested that by using a computer to analyze a symbol-based piece of music (i.e. MIDI), a certain mood could be derived.[14] By analyzing aspects of music such as timbre, tonality, tempo, and rhythm, a specific mood could be guessed by a computer. By pairing this with a system that takes user feedback and trains itself, the system now becomes much more accurate and has produced astonishing results.

[13] (De Silva, Ranganath, & De Silva, 2008)
[14] (Wang, Hang, & Zhu, 2004)

It has been proposed that music instills a specific set of emotions and that these emotions can be detected using computers. Robert Thayer presented a model called 2-dimensional emotion space (2DES)[15] in which emotion is categorized on two axes, arousal and valence. An example of 2DES is shown in Figure 7. It shows on two axes what emotions are present for the different possibilities of music. Arousal can be either energetic or silent while valence can be either positive or negative. A computer can calculate arousal by analyzing factors such as tempo, volume, and note density. However, human ears more easily detect other things such as key and lyrics.

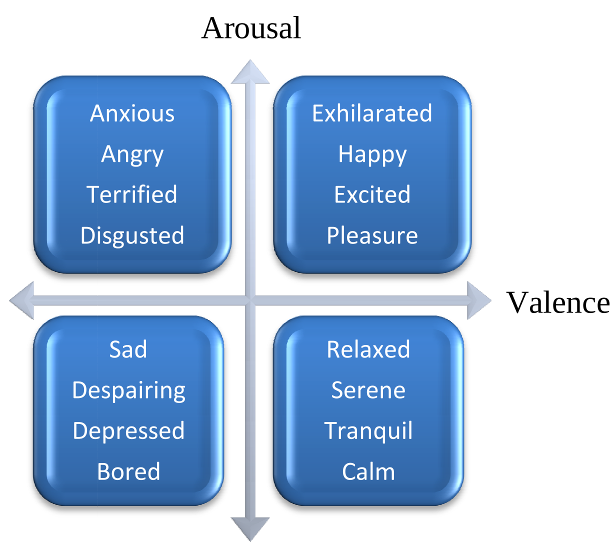


Figure 7 - 2DES Model

Thayer's approach to detecting mood in music was limited to MIDI music, where notes, instrument, and other variables are readily available. Research by Yang, Liu, Wu, and

[15] (Thayer, 1989)

Chen[16][17] has adapted the 2DES model but has applied it to popular music. Rather than analyzing MIDI files, their system analyzes music in WAV format, extracting features such as volume, tempo, and note density. Also, rather than trying to classify a specific mood or emotion, they use fuzzy logic to determine which quadrant of the 2DES model a specific song might fall into.

4.7 Tracking Eye Movements

Another facial feature that could be used is the movement of the eyes. This concept was explored in another project advised by Professor Bianchi called “Human Bandwidth Optimization Methodology”. It was suggested in their project that by tracking eye movements, a television might know when you have looked away from it. When looking at a webpage, you might be able to simply look at the down arrow to scroll down a page. Also suggested was the fact that an entirely new “eye-movement language” could be created.[18]

Andrea Polli developed a system that can track her eyes and do various things with the data received.[19] By using a grid of nine words, she was able to create a digital piece of work that would playback the words she looked at without using her hands. Figure 5 shows her performing with a flutist wearing her device as a visor. She reports that the use of the eyes as a method of digital input can be very different than one would think. You can easily look up or down whenever you want. However, there are many involuntary

[16] (Yang, Liu, & Chen, Music Emotion Classification: A Fuzzy Approach, 2006)
[17] (Yang, Liu, Wu, & Chen, 2006)
[18] (Knapp & Runes, 2005)
[19] (Polli, 2001)

movements made, especially when playing music. Nevertheless, her system is a breakthrough in expanding human-computer bandwidth.

Using movements similar to those made by the fingers on the surface of Apple's new iPhone, such as the flick or drag, perhaps eye movements could be interpreted to scroll up or down a page. A blink could mean click while a double blink would double click the icon you are looking at. Gestures like these have been put to use with a system for physically challenged patients. By using eye movements, test subjects were able to input words into a computer simply by looking at a letter and "double-blinking".[20] Borgetti, Bruni, Fabbrini, and Sartucci also suggest that this system can easily be updated to support other functions such as control games, input other types of symbols, text-to-speech, and even controlling a motorized wheelchair. With our current technology, we will soon be able to completely do away with our mice and keyboards because of things like Voice Recognition and eye-movement tracking.

Figure 8 - Polli performing using Intuitive Ocusonic system (Polli, 2001)

[20] (Borgetti, Bruni, Fabbrini, & Sartucci, 2007)

4.8 Augmented Reality

Most of us are familiar with the idea of virtual reality. A predefined environment is presented in a 3-Dimensional virtual world. By using a specific type of controller and visor, you can walk around and manipulate objects in the virtual world. There is a new spin on virtual reality that is being looked into, and this is the idea of "Augmented Reality".[21]

The idea behind augmented reality is that rather than having a virtual world that has been created by a programmer, the world that the user would see is the real world. Through the use of multiple cameras, head motion tracking, and a sort of 3-D tracking device, people would be able to explore the world around them. This process is more complicated than it seems. If you are to explore the real world in a virtual fashion, many things need to be addressed. With a preprogrammed virtual world, there are defined objects with defined dimensions that can be manipulated. If we wanted more than one user to be able to focus on a certain object, it would be very simple to just get the virtual coordinates and focus everybody's view on it. Augmented reality is very different. It is dealing with multiple cameras and objects that have not been defined by a computer program.

One simple example of augmented reality is the use of the yellow line in football that shows where the first down marker is. Anybody not watching on TV is unable to see this line. It is an automated line created by a computer. This computer knows where every yard is on the field and can determine where to place the line depending on the teams' current position. The main task in augmented reality is object detection. An image cannot just be stationary in the user's field of view (like the lines manually drawn on the screen by commentators to explain a play in football). It needs to be dependent on the

[21] (Siegla, Hanheideb, Wredeb, & Pinz, 2007)

environment, requiring that a computer be able to assign coordinates in the real world and place a virtual object within the virtual world.

Imagine being at home and being able to go to the grocery store and not even have to leave your living room. Maybe you are in the library and cannot read certain titles on the top shelf. You can put on your visor and fly up to the top shelf to read any title you wish.

Again, other than just having a fun use, soldiers might be able to explore the battlefield simply by setting up cameras in certain strategic locations. Maybe a remote control device could position multiple cameras, allowing the soldier to walk through this "augmented reality" without the threat of danger.

5 Proposed Ideas

Using the methods of expanding bandwidth above, it is evident that technology is getting more sophisticated. Although research on the topics is still underway and far from over, we can still get an idea of what type of technology might be available within the next several years. This section will discuss what some of the new technologies might have in store for the future of human bandwidth.

5.1 Emotion-Based Digital Music

With technology like emotion detection in music, it will not be long before MP3 files are tagged with an additional “Mood” field. Additionally, with advances in facial expression detection, computers will easily be able to determine what type of mood somebody is in. Combined, these two technologies can change the way people listen to music on their computer.

Depending on the type of music somebody enjoys, a computer will know the mood of all songs in a person’s library. When a person sits down to use the computer, the computer will detect the mood that person is in using facial expression detection. If somebody is very relaxed, perhaps the computer will detect that and begin to play a song that is positive in valence (i.e. Major key and positive lyrics). The user might even have control over factors such as increasing or decreasing the arousal of songs being played. While in a relaxed state, somebody might want to become more energetic, cueing the computer to begin increasing the level of arousal in the music while remaining at a positive valence.

One website that has already begun a project in determining the “feel” of a song is Pandora (http://www.pandora.com). This website starts by asking for a song that the user enjoys. Given the mode, lyrics, instrumentation, and key, it will play similar music. Combined with user input, Pandora will begin to get a better idea of what type of music any particular person likes.

Today's technology and research is at a point where mood detection in popular music is accurate. Additionally, the accuracy of facial expression detection is very high as well. I believe that if somebody were to begin tagging music with mood, it would be very easy for a computer to play music depending on that mood. With other types of algorithms and more data, the computer might even be able to play music that will change a person's mood.

Pandora is just an example of how listening to music on your computer is beginning to change. With newer technology like emotion detection and facial expression detection, perhaps Pandora in the near future will begin playing music without even asking for any more user input other than just sitting in front of the computer.

5.2 Uses for Augmented Reality

Games such as World of Warcraft, Battlefield, and Halo put the player in the middle of a virtual world, allowing them to walk around, explore, fight, and interact with objects. This is an example of virtual reality. The game usually ends when you turn off your computer. With augmented reality, it may be possible to bring the game with you. A great example would use augmented reality with a MMO such as World of Warcraft to create an environment in the real world. Users would need a special "see through" visor display, allowing virtual objects and information to be displayed in real time on the actual surroundings. If somebody else around you is playing, the visor would alert you and you would then have the option to go and start a virtual battle with that person.

A more realistic use for augmented reality would be to have soldiers use it when they are in the field. Similar to a Heads-Up Display (HUD) in an aircraft cockpit, soldiers would be able to simply wear a visor that had information such as heading, temperature, wind direction and speed, ammo left, and the positions of all squad members.

Simple data such as wind speed, wind direction, remaining ammunition, and temperature is very simple to display. Research using GPS along with augmented reality opens the door for countless possibilities. A soldier might simply be able to look around in the environment and have his objective building highlighted by a green outline, telling him exactly how far away it is, how many people are estimated to be inside, and even highlight possible methods of entry.

In order to make this possible, many of the new technologies mentioned earlier would need to be able to work flawlessly and also work with each other. GPS would be able to read and process the user's absolute position in the real world. GPS would even be able to detect the individual's speed. Different tracking devices would need to be worn on the head, specifically head motion and eye tracking equipment. In order to tell the display exactly what to show and from what perspective, the system needs to know which way a person is looking, whether they turn their head or just look with their eyes. Finally, augmented reality would be responsible for looking at objects in the real world, defining them in an augmented reality, and displaying them to the user as a virtual object. This would allow the user to look at a distant building, receive important information, and possibly even manipulate their view on the building.

5.3 Composing Music with the Mind

For any composer who wishes to create a song in their heads and put it into a digital format, the easiest way to do this would be to think of the song and have it recorded. This idea is obviously far from existence. The reality is that newer technology shows reading a mind might just be as simple as recording and interpreting subvocal speech.

Using subvocalization might be simpler than some people are making it. Rather than focusing on specific words and sounds made by somebody and trying to recreate actual words, perhaps using subvocalization on a more qualitative scale would prove to be useful also. For example, perhaps a computer could guess at least the rhythm of a song by detecting pulses in subvocalization.

6 Conclusion

Human bandwidth is very wide and the number of ways people are able to interact are countless. Between two people, the interaction is usually flawless. Subconscious signals can be interpreted and facial expressions easily analyzed and used. The drawbacks come when people are asked to interact with something that is not another human. Interaction with computers has been limited to a keyboard, mouse, monitor, and speakers. Musicians are limited to MIDI controllers such as a MIDI keyboard, mouse and computer keyboard, and even alternate methods using scanners and gesture-based input systems.

Much research is going into expansion of human bandwidth, and some of the options have been discussed in this paper. Researchers have known about subvocalization for many years, however only recently has NASA been able to actually detect and interpret subvocal signals. In addition to subvocalization, the methods of detecting subconscious actions used by computers have been expanded greatly. Technologies such as facial expression detection and eye tracking are only two examples.

In order to put these things to use, computers need alternate methods of outputting data to people. Emotion detection for digital pieces of music adds a new way of allowing users to choose their music. Rather than by artist, title, or genre, people will soon have the option to choose music by mood or emotion. Paired with facial expression detection, people may not even need to choose their music, but instead sit down at their computer and just hit a "Play Music" button.

Finally, augmented reality is beginning to take hold in everyday life. Yellow yard-markers in football and HUDs in airplanes and some cars are only two examples of how augmented reality is already being used. If paired with eye and head movement tracking, a see through head mounted display, and object detection, augmented reality may be the next big thing. Soldiers can use it in combat situations, people can use it while walking or driving down the street, and even surgeons can use it when performing surgery. It may also take over the video game world by allowing gamers to play video games based in the real world.

7 Bibliography

Anstice, J., Bell, T., Cockburn, A., & Setchell, M. (1996). The Design of a Pen-Based Musical Input System. *6th Australian Conference on Computer-Human Interaction* (p. 0260). OZCHI.

Bitar. (n.d.). Instrumentation Amplifier. *ECE 2011 Lab Exercises* .

Borgetti, D., Bruni, A., Fabbrini, M., & Sartucci, F. (2007). A low-cost interface for control of computer functions by means of eye movements. *Computers in Biology and Medicine* , Borghetti D., Bruni A., Fabbrini M., Murri L. and Sartucci F.

Brewvana. (n.d.). Retrieved from Brewvana: http://brewvana.wordpress.com/category/music/

Carver. (1990). *Wikipedia*. Retrieved September 2007, from Wikipedia: http://en.wikipedia.org/wiki/Subvocalization

De Silva, C. R., Ranganath, S., & De Silva, L. C. (2008). Cloud basis function neural network: A modified RBF network architecture for holistic facial expression recognition. *Pattern Recognition , 41* (4), 1241.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* , 124-129.

Freud, S. (1963). *Dora: An Analysis of a Case of Hysteria.*

Hamman, M. (2000). Priming computer-assisted music composition through design of human/computer interaction. *Mathematics and Computers in Modern Science - Acoustics and Music, Biology and Chemistry, Business and Economics, Mathematics and Computers in Modern Science - Acoustics and Music, Biology and Chemistry, Business and Economics* , 75-82.

Hardyck, C. D., Petrinovich, L. F., & Ellsworth, D. W. (1966, December 16). Feedback of Speech Muscle Activity during Silent Reading: Rapid Extinction. *Science* , pp. 1467-1468.

Hart, D. (n.d.). NASA Ames Research Center.

Jorgensen, C., & Binstead, K. (2005). Web browser control using EMG based sub vocal speech recognition. *38th Annual Hawaii International Conference on System Sciences* (p. 294). Big Island, HI: IEEE Computer Society.

Jorgensen, C., Lee, D., & Agabon, S. (2003). Sub Auditory Speech Recognition Based on EMG Signals. *International Joint Conference on Neural Networks* (pp. 3128-3133). IEEE Computer Society.

Kirlik, A., & Maruyama, S. (2004). Human-technology interaction and music perception and performance: Toward the robust design of sociotechnical systems. *Proceedings of the IEEE, v 92, n 4, April, 2004, Engineering and Music* (pp. 616-631). IEEE.

Knapp, R., & Runes, J. (2005, July 24). Human Bandwidth Optimization Methodology.

Locke, J. L., & Fehr, F. S. (1972). Subvocalization of Heard or Seen Words Prior to Spoken or Written Recall. *The American Journal of Psychology* , 63-68.

Miyao, H., & Maruyama, M. (2007). An online handwritten music symbol recognition system. *International Journal on Document Analysis and Recognition , 9* (1), 49-58.

Ng, E., Bell, T., & Cockburn, A. (1998). Improvements to a Pen-Based Musical Input System. *Computer Human Interaction Conference* (pp. 178-185). Adelaide, SA, Australia: IEEE.

Pauws, S., Bouwhuis, D., & Eggen, B. (2000). Programming and enjoying music with your eyes closed. *SIGCHI conference on Human factors in computing systems* (pp. 376-383). The Hague, Netherlands: ACM, New York, NY, USA.

Peretz, I., & Zatorre, R. J. (2005). BRAIN ORGANIZATION FOR MUSIC PROCESSING. *Annual Review of Psychology* , 89.

Polli, A. (2001). Rapid Fire: Eye Movements in Human Computer Interfaces. *Convergence* , 36-46.

Siegla, H., Hanheideb, M., Wredeb, S., & Pinz, A. (2007). An augmented reality humancomputer interface for object localization in a cognitive vision system. *Image and Vision Computing* , 1895-1903.

Thayer, R. (1989). *The Biopsychology of Mood and Arousal.* New York: Oxford University Press.

Wanderley, M. M., & Orio, N. (2002). Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal* , 62-76.

Wang, M., Hang, N., & Zhu, H. (2004). User-adaptive music emotion recognition. *2004 7th International Conference on Signal Processing Proceedings* (pp. 1352-1355). Beijing: IEEE.

Wikipedia. (2007). Retrieved from Wikipedia: http://upload.wikimedia.org/wikipedia/commons/d/d4/Illu01_head_neck.jpg