Composing Music using Vocal Synthesis and Other Vocal Effects

Myles Karna
Northeastern University
Boston, United States
karna.m@northeastern.edu

ABSTRACT

Music has a unique quality to it in that it brings together vast groups of people using the power of sonic space. It has evolved drastically over the course of the past half century technologically, and one of the most fascinating pieces of technologies is vocal synthesis, which manipulates the voice in the form of an incoming signal in various ways. This paper showcases the use of different types of vocal synthesis in the form of an original song, accompanied by various forms of instrumentation that contribute to the overall sonic tone of the piece.

1. INTRODUCTION

Music, and sound in general, has the power to elicit a multitude of different potent emotions through the use of different chords, tones, timbres, and other musical features that contribute to the creation of a uniquely curated sound environment. One of the most unique ways of creating music is through vocal synthesis and other vocal effects and techniques.

1.1 History of the Vocoder

Over the last 50 years, musical pieces have evolved drastically, as the technological advances in music have coincided with the advance of the general technology of the world at the time. Effects such as distortion and reverb, along with others were not discovered or invented until around the 1970s onward. One effect that took the world by storm was vocal synthesis, which at first manifested itself as a sizable apparatus that was only accessible by a select few. A popular offshoot of vocal synthesis is the vocoder, which takes the incoming signal, in this case a human voice, and manipulates it in a way that allows for the voice to be played as an instrument on the piano. Various chords can be played in unison with sung or spoken voice, and the output is the human voice, but heavily altered into the chordal structure that is being played. The

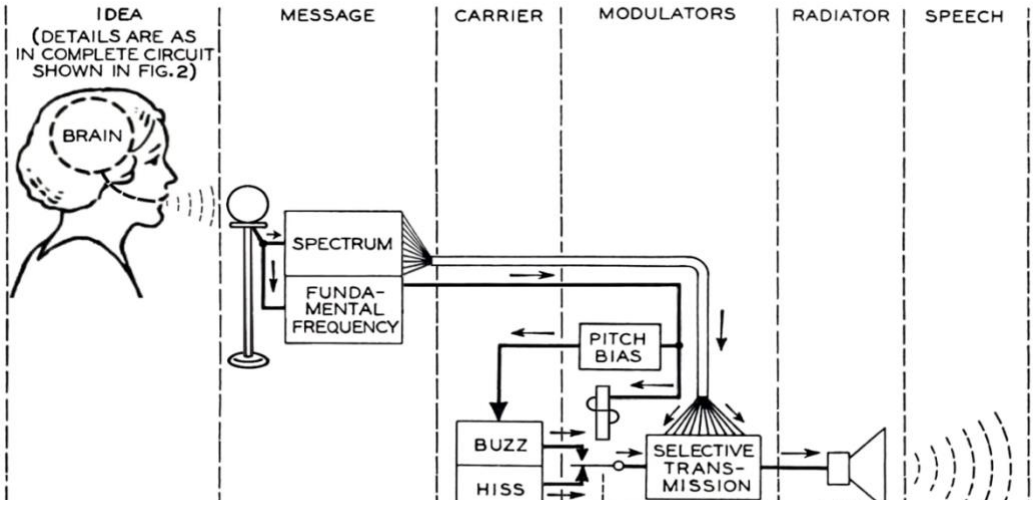


Figure 1: Homer Dudley's circuit diagram of his first vocoder

vocoder was first developed by Bell Labs engineer and physicist Homer Dudley in 1928 [4, 5]. The machine consisted of two parts: an analyzer and a carrier (or synthesizer). The analyzer received incoming auditory signals and split it into its separate frequencies and overtones on a spectral graph. The synthesizer essentially reversed the process by inputting these frequencies into auditory filters and oscillators that recreated the initial input signal. The first vocoder was created simply in an attempt to transmit telephone communications with a reduced bandwidth [4]. It wasn't until the 1970s that vocoders began their life as a tool for music, with the likes of Boney M. and other notable artists making use of it in their songs. One of the first appearances of the vocoder in music was when Bob Moog, the creator of the Moog synthesizer, and composer Wendy Carlos used the vocoder in the soundtrack to *A Clockwork Orange* in 1971. Phil Collins's hit, "In The Air Tonight" famously utilizes the vocoder in the background of the verses of the song. One of the most recognizable instances of the vocoder is in Kraftwerk's song "The Robots," in which the vocoder is the lead vocal part. Kraftwerk were considered one of the pioneers of the use of vocoders in mainstream music, as the vocoder fit perfectly into their style of robotic electronic music.

Nowadays, the technology involved in vocal synthesis has become exponentially more complex, yet increasingly accessible, to the point where anybody can access the technology on their computer or digital audio workstation. The specific software in use for the piece of music highlighted in this paper is called VocalSynth 2, developed by Izotope in 2018.

Vocoders and voders have also had extensive use outside of the realm of music. As previously stated, the initial purpose for the voder was to assist in telecommunication. Nowadays, there are other ways in which this technology is utilized, one of which is speech synthesis for those who are and/or hearing impaired.

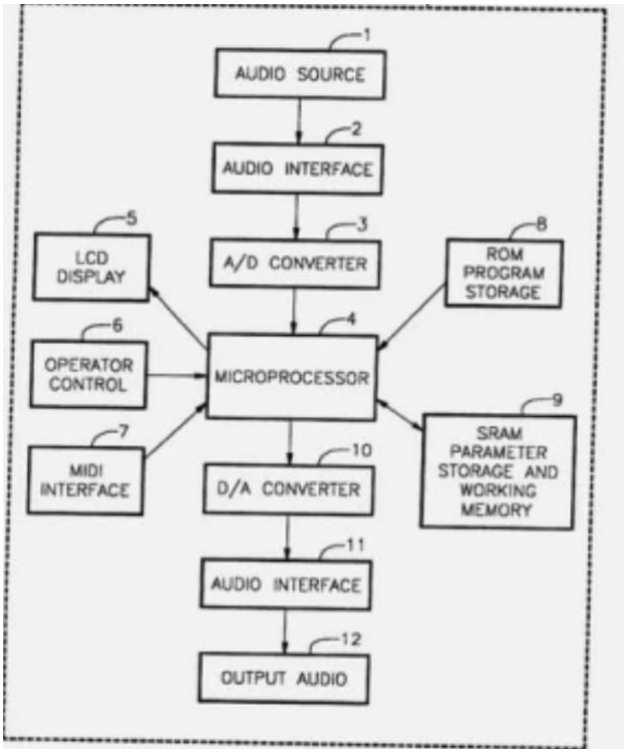
1.2 History of Autotune

Another vocal effect used in this piece is autotune, which has come under scrutiny ever since its invention in 1996. Autotune was created by Dr. Andy Hildebrand, a former oil engineer, after a conversation he had with the wife of a colleague over lunch, who, when asked by Hildebrand "What needs to be invented," responds with "Why won't you make me a box that will let me sing in tune? [2]" Hildebrand initially dismissed the question as a joke, but after returning to the question later, he realized that this may be possible. At this time, the software behind pitch correction did not exist, and was viewed as the ultimate goal to reach by many music inventors.

Autotune, as it was first created, works by applying autocorrelation to the incoming audio signal. Hildebrand, as a former oil engineer, employed the use of autocorrelation in his everyday life to process and read projected seismic signals from underground oil. He realized quickly that he could employ the same techniques to music:

> *"When you're processing pitch, you add wave cycles to go sharp, and subtract them when you go flat. With autocorrelation, you have a clearly identifiable event that tells you what the period of repetition for repeated peak values is. It's never fooled by the changing waveform. It's very elegant* [2]."

The processing power required for this amount of computation was largely unachievable at this time, but Hildebrand discovered a way to reduce the amount of computation needed by consolidating "a million multiply adds into just four [2]." This mathematical trick saved Hildebrand years of computation and algorithms and enabled him to develop Autotune right away in early 1996.



Hildebrand brought his new invention to the National Association of Music Merchants (NAMM) conference later that year, where he became the most popular person at the conference immediately. Hildebrand had become the first person to successfully create a tool to attain pitch-perfect vocals.

Initially, Autotune was kept a secret that only some recording studios had access to. However, once Cher's hit song "Believe" was released, the music industry changed forever. "Believe" wasn't the first song to make use of Autotune, but the first to make use of it in a way that was made obvious to even the least informed listener. In between Autotune's formal release and "Believe" being released, artists and producers had used Autotune sparingly to correct minor mistakes in pitch, while maintaining their natural voice. Cher's "Believe" utilized Autotune in such a way that altered her voice. "Believe" sold over 10 million copies and alerted the world of this new technology. Artists everywhere perked up their ears and wanted to try it. One of those artists was T-Pain, who ended up being one of the pioneers of Autotune, for better and for worse. T-Pain made his start in 2004 after he was signed to Akon's record label Konvict Muzik. He initially started out in the rap sphere, but realized fairly quickly that he would rather sing. In his hometown of Tallahassee, Florida, according to him, "I wanted to be a singer instead of a rapper because everybody in Tallahassee was rapping. And if I wanted to get out of Tallahasee and I wanted to make something of myself, I had to do something different [1]." It took two years for T-Pain to discover what could help him stand out, and that came in the form of Autotune.

At this same time, however, Autotune started garnering increasing amounts of criticism from the public, with many touting it as a "cheat" and that every artist who used it couldn't sing. Artists such as Jay-Z and Usher openly denounced the use of Autotune.

Nowadays, almost every major artist in the world uses Autotune to some degree, whether they're using it sparingly or like T-Pain and significantly altering their voice. Despite the backlash that Autotune has faced, it has proved to be an invaluable tool in the music industry for achieving clear vocals.

2. VOCALSYNTH 2

VocalSynth is an immensely powerful software plugin that incorporates five separate vocal synthesis effects into one encompassing plugin:

1. Biovox
2. Vocoder
3. Compuvox
4. Talkbox
5. Polyvox.

Each of these effects outputs its own unique sound based on the incoming signal, and each of the five effects can be mixed and matched to further increase the number of distinct tones and timbres that can be produced. The vocoder and talkbox effects are the original effects (or at least

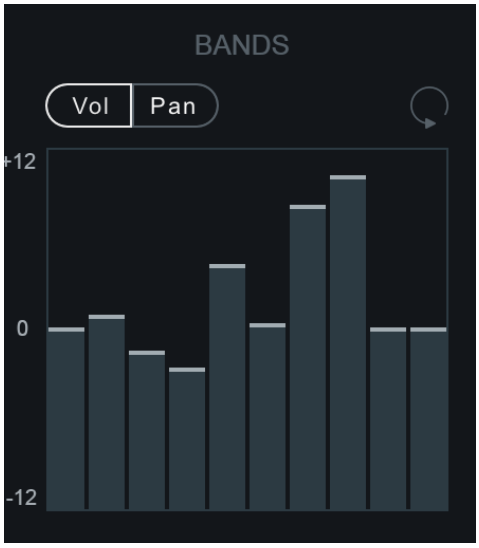
similar copies) that were initially created many years ago.

In regard to the other three effects in VocalSynth, Biovox aims to emulate the human voice in its vocal processing and effect output. It contains four different knobs that control the human aspect of the effect: Clarity, Shift, Breath, and Nasal. The clarity knob does exactly what the name entails: it adds more clarity to the vocal by mixing in rawer signal. It also incorporates vocal formant shifting in the form of vowel manipulation. The Breath and Nasal knobs are also self-explanatory, adding more breathiness and nasality, respectively, to the outgoing signal.

The Compuvox effect is a unique effect that incorporates bit crushing distortion into the outgoing signal, which outputs a very unique type of grainy vocal effect. With this effect, the user can add more or less bytes, which corresponds with more or less distortion, respectively. The Polyvox effect is a polyphonic pitch shifter that also attempts to emulate the human voice, in the form of lush harmonies that resemble a human choir. The user can tune the pitch of the incoming signal to different notes, which outputs a chord, either major or minor, of the user's choosing.

With this plug-in involving vocal synthesis, each vocal effect within VocalSynth has a box dedicated to the manipulation of the base synthesizer that the effects are following. In this box, one can change the type of oscillator that is the input signal follows, whether it be a sawtooth wave, triangle wave, or square wave. Another oscillator can be added on top of the first oscillator, with the same options for wave type being available. Along with the two oscillators, there is a noise button that adds white noise to the overall signal. This can have multiple uses, such as adding more space, or adding more grit to the tone of the synthesizer. In the Biovox effect, along with the previously stated optons, there is also a graph with different vowels on it and a slider, which can be moved along the graph to emphasize different vowel sounds. This is extremely valuable when placing the vocal effect onto a specific frequency, as different vowel sounds emphasize different frequency bands. For example, the "oo" and "oh" sounds place more emphasis on lower frequencies, whereas "ee" and "i" accentuate higher frequencies.

In the Talkbox effect, there is a similar graph to the one in the Biovox effect, except this graph is a linear column graph, in which each column represents a specific frequency band. Increasing these columns increases the specific frequency, and decreasing them decreases the specific frequency, similar to a basic EQ effect. This has a very similar outcome to the vowel graph of the Biovox.



There are three different modes that can be used in VocalSynth: Auto, which automatically outputs a signal based of the note being sung in real time, MIDI, which takes in MIDI input notes on which the plugin maps the voice to, and Sidechain, which allows the user to use different instruments that can be run through VocalSynth to achieve unique sounds. For this original song, all of the relevant vocal performances have been manipulated using the MIDI function.

3. THE SONG

3.1 The Meaning Behind the Song

The song for this project is an original song entitled "Seasons" that was specifically written for this endeavor. It is a song that discusses the effects of Seasonal Affective Disorder, or Seasonal Depression, on the author. It contrasts the beauty of the environment and the scenery of the outside world during the winter months with the dangerous effects that this time of year brings resulting from lower amounts of sunlight per day, paired with the increased overall cloudy atmosphere at this time of year. Seasonal Affective Disorder is a topic that the author is very familiar with, having struggled with it for many years.

3.2 Song Structure

"Seasons" is in verse-chorus form with an introductory verse at the beginning and is in the key of E Major. The chord structure in the verses follows a basic I-IV alternation in the first half of each phrase, and switches to a vi-V-IV descension in the second half. In the chorus, the basic chord structure is I-V-vi-V-IV, but the chords played by the synthesizer are more complex. The E Major chord transitions through a C#7sus4 to the V, which, in the key of E, is B Major. As the songwriter, I made this decision based on the lyrics that are sung at the time: "What have I become?" This question, which opens the chorus, conveys the feeling of being hopeless and lost. I felt that the pang of the dominant 7^{th} chord substitution perfectly accompanied that feeling. The vi-V-IV walk down coincides with the lyrics "I'm broken and bleeding / I'm picking up pieces now," and "I'm down fatigued all because of the seasons," which emphasize the idea of feeling depressed, lost, and unsure of how to move forward.

3.3 VocalSynth in the Song

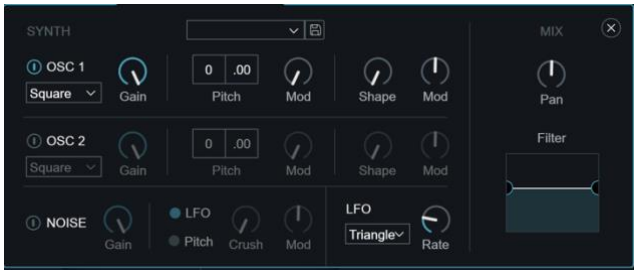
In regard to what makes up the song sonically, there are a couple of tracks that play a significant role in the sonic space. There are three separate vocal tracks that serve as both lead vocals and as pointers to three respective MIDI tracks that control VocalSynth's vocal synthesizer.

The first synthesizer is paired with the intro and verse vocals, as well the final chorus vocals. This synthesizer makes use of the BioVox effect, as well as the Vocoder effect. BioVox has remained relatively unchanged, with the pointer in the center of the vowel graph. The clarity knob has been set to the maximum, as this portion of the vocal synthesizer sounded muddy and unclear. A bit of nasality has been added as well to enhance the human aspect of this portion of the vocal synthesizer. The Vocoder effect has been heavily altered, with most of the work going into the manipulation of the frequency bands on this

vocal track. A heavy emphasis has been put on the higher frequency bands to further enhance clarity, and the lower frequency bands have been lowered in order to both lighten the intro and first verse, as well as leave sonic headspace for lower-frequency accompaniment, like piano, guitar, and bass, later on in the piece.

The second vocal synthesizer controls the first chorus. It's very similar in tone to the first synth, but there is much more emphasis on lower frequencies now that the introduction is over. This is done by introducing the Compuvox effect, which has been restricted to the low frequencies and serves as the bass of this synthesizer. I have also added slight ring modulation to achieve a unique vibrato effect, as well as a sweeping filter at the end of the chorus for flare. Finally, I added an EQ to shape the high-end frequencies.

The third and final synthesizer controls the pre-chorus that follows the first verse, as well as the second verse. It has a very distinct tone that is different to the other two synthesizers, and that is due to the use of the Talkbox and Compuvox. The Compuvox's oscillator is a square shape and is set to a very smooth, high-bit setting, whereas the Talkbox's oscillator is the grittier sawtooth wave and is set with a very high drive, which adds distortion and grittiness to the signal. These two effects, when paired, creates a very unique type of distorted voice signal. This sonic decision was made based on the emotions conveyed by the lyrics sung on top of the synth. In the pre-chorus, the character's desperation and hopelessness is reaching a peak, which, lyrically, points to a lack of feeling human. The same can be said for the second verse. Thus, the decision was made to emphasize the robotic nature of the vocal synthesizer, to the point where it almost drowns out the lead vocal.



3.4 Autotune in the Song

Autotune plays a critical role stylistically in this song, and is present on all three of the main vocal tracks. The decision to use Autotune in this song was stylistic, and it plays a role in contrasting the very human and sensitive lyrics in the song. Autotune has two main knobs:

1. Retune Speed
2. Humanize

Retune speed controls the speed at which the incoming pitch signal is corrected, from slow to fast. To achieve a more natural and subtle correction, slow correction is best, and for a more noticeable correction, fast correction is best. The other knob, Humanize, sustains the ending of each note to, as the name states, "humanize" the vocals. On this song, to enhance the artificial nature of the vocals, the retune speed is set on the fastest setting, and the humanize knob is set to 0.

4. CONCLUSION

The purpose of this piece was to demonstrate the use and purpose of vocoders and autotune, and to show that they can be utilized in a way that doesn't involve "cheating" or "faking," but rather as a stylistic choice. The use of vocoders and Autotune have helped pave the way for increased creativity and uniqueness in composing and songwriting, and has helped open new avenues in the creation of music. In this piece of music, the purpose and role of VocalSynth and Autotune, both as tools, and as markers for change have been showcased in a way that allows them to stand out in the song. Both of these tools have a decorated past, whether that be through the creation and improvement of telecommunication, in the case of vocoders, or through a breakthrough in math and science against all odds, in the case of Autotune.

REFERENCES

[1] M. Evstatieva, "T-Pain Rises Above Haters," NPR, https://www.npr.org/2014/11/13/363759981/t-pain-rises-above-the-haters (accessed Apr. 22, 2024).

[2] Z. Crockett, "The mathematical genius of auto-tune," Priceonomics, https://priceonomics.com/the-inventor-of-auto-tune/ (accessed Apr. 22, 2024).

[3] "Auto-tune - the best vocal plug-ins available," Auto-Tune - The Best Vocal Plug-Ins Available, https://www.antarestech.com/community/vocoder-history (accessed Apr. 12, 2024).

[4] M. McAllister, "A short history of the vocoder," Produce Like A Pro, https://producelikeapro.com/blog/a-short-history-of-the-vocoder/ (accessed Apr. 12, 2024).

[5] "A brief history of the vocoder," iZotope, https://www.izotope.com/en/learn/a-brief-history-of-the-vocoder.html (accessed Apr. 12, 2024).