VOCALOID: Fandom's Influence on Professional Tools

Jay Verchin

MUSC 127 Final

Spring 2019

Prof. *Tina Tallon*

Overview

The Vocaloid engine, a singing synthesis tool, was released by Yamaha Corporation with the virtual singer libraries LEON and LOLA (Fig. 1) on January 15, 2004. The product was initially marketed to professional musicians as a replacement for a human voice. However, the product had trouble taking off at first and didn't see major success until around 2008, with the release of digital diva, Hatsune Miku for the Vocaloid 2 engine. From there, Vocaloid gathered a large fandom and saw international success, releasing five different versions of the engine and generating a large fanbase ranging from teenagers to professional musicians.

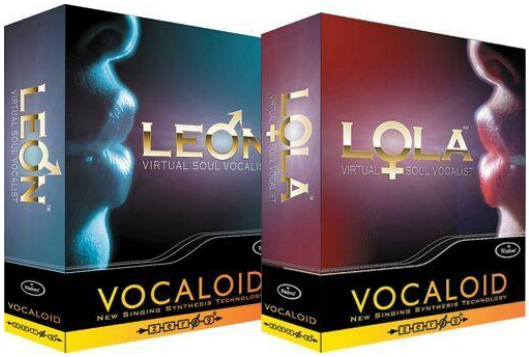

Figure 1: Vocaloid LEON and LOLA's box-art.[1]

A Brief History of Vocaloid

The original Vocaloid engine was released in 2004, however its software began development in 2000. The product began development in Japan where Kenmochi Hideki proposed Vocaloid's concept. The technology for the software was then researched at the Pomeu Fabra University in Spain.[2] At first, this research was purely academic curiosity, and was never intended for commercial use. However, Vocaloid was regarded as a trailblazer in tackling modern synthesis of singing Vocals, so studios Crypton Future Media and Zero-G collaborated with Kenmochi and Yamaha to make the software a widely available product. Five original Vocaloid products were released in both Japanese and English versions of the program. The three English Voicebanks were named LEON, LOLA, and MIRIAM and the

two Japanese Vocaloids were named MEIKO and KAITO. All three English Vocaloids were marketed with stock-image box-art, making them marketable to professional musicians. LEON and LOLA's voice providers remain a mystery, part of Miriam's appeal is idea that "the user may be offered access to the familiar face and voice of a popular singer,"[3] which was British Singer Miriam Stockley, who Yamaha and Zero-G hoped would help sell their product. Crypton Future Media tried a similar approach with one of their two voicebanks, MEIKO, who was voiced by Meiko Haigou, a popular Japanese vocalist.

Unfortunately, the original Vocaloid engine wasn't popular with many, including the professionals the product was intended for. That all changed with the release of Hatsune Miku (Fig. 2) on August 31st, 2007, when Crypton Future Media took a different approach to their first product for the Vocaloid 2 engine. This was the first time that character design was considered before voice quality, and it revolutionized the Vocaloid industry. Miku was marketed as a virtual idol with an unforgettable green-twintail look. Unlike the first set of Vocaloid libraries that were voiced by professional singers, Miku's voice provider was Saki Fujita,[4] a Japanese voice actress that gave Miku a cute, sweet sounding voice to go with her look. This made her widely appealing to anime fans which boosted the Vocaloid 2 engine's sales and put Vocaloid globally on the map. She saw an "append" release for the Vocaloid 2 software with six additional voicebanks varying in tone[5] and re-releases for the Vocaloid 3 and 4 engines. She also has appeared on David Letterman's Late Show, in a

Figure 2: Vocaloid Hatsune Miku's box-art.[4]

Japanese shampoo commercial with Scarlett Johansson, and on Lady Gaga's ARTPOP tour as her opening act, establishing the virtual diva's place as an international pop icon.

Crypton Future Media's success translated into more voice libraries, such as Kagamine Rin and Len (2007, re-release in 2008) as twin Japanese banks, and Megurine Luka (2009) as the first bilingual voice with support for both Japanese and English.[6] These "Cryptonloids" dominated the market over voices such as voices like Prima and Tonio, for example, which were opera style singers developed by Zero-G. This emphasis on character over voice quality marked a clear shift in the target group of the market from professional musicians to anime fans and a growing Vocaloid Fandom.

So, How Does Miku Sing?

The Vocaloid engine, from the first version in 2004 to the latest version released in 2018, is a concatenative synthesis-based program based on a source-filter model.[7] This allows pre-recorded phonemes are recorded and played back in a certain order and at certain pitches to create melodies using a sound library as the sample sources. (Fig. 3) For the Vocaloid and Vocaloid 2 engines' Japan ese voices, this was done using a very simplistic form of diphone consonant-vowel recording, made possible by the simplistic syllables of the Japanese language. To understand the recording patterns of Japanese Vocaloid voicebanks

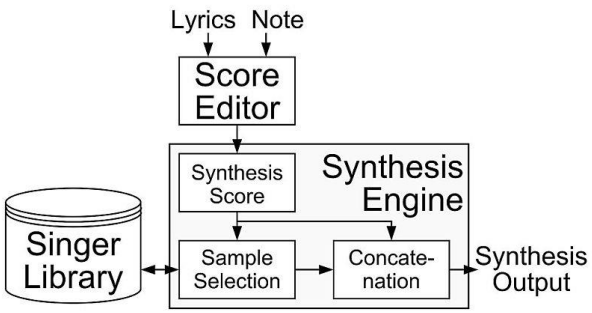

Figure 3: Schematic of the Vocaloid Engine's concatenative synthesis system.[7]

in more depth, it is important to understand a little bit about linguistics of the Japanese language.

Hiragana (ひらがな)											
n	wa	ra	ya	ma	ha	na	ta	sa	ka	a	
ん n	わ wa	ら ra	や ya	ま ma	は ha	な na	た ta	さ sa	か ka	あ a	a
		り ri		み mi	ひ hi	に ni	ち chi	し shi	き ki	い i	i
		る ru	ゆ yu	む mu	ふ fu	ぬ nu	つ tsu	す su	く ku	う u	u
		れ re		め me	へ he	ね ne	て te	せ se	け ke	え e	e
	を wo	ろ ro	よ yo	も mo	ほ ho	の no	と to	そ so	こ ko	お o	o

Table 1: Phonetic character chart for Hiragana, one of the Japanese alphabets. [8]

As seen in this hiragana chart, there are only five vowels in the Japanese language, denoted in IPA as [a], [i], [ɯ], [e], and [o].[9] There is also a prolonged "n" sound, denoted by the IPA [ɴ][9] that takes up the same amount of time as and functions similarly to any other vowel, though no other consonants precede it. Japanese only has about 15 consonants as well,[10] making the number of phonemes necessary for a voice library significantly less than that of English.

This allows for the easy separation of syllables, allowing for a diphone "consonant-vowel" system. However, this approach leaves a large separation between syllables (Fig. 4), making the vocals sound choppy and robotic. This problem was remedied in the Vocaloid 3, 4, and 5 engines with triphone support. This recording style chains syllables together, allowing the samples to be configured into demi-syllables in a vowel-consonant-vowel

format. After that, the syllables can be crossfaded together (Fig. 5) to create a continuous sound that is less choppy than consonant-vowel diphone systems.

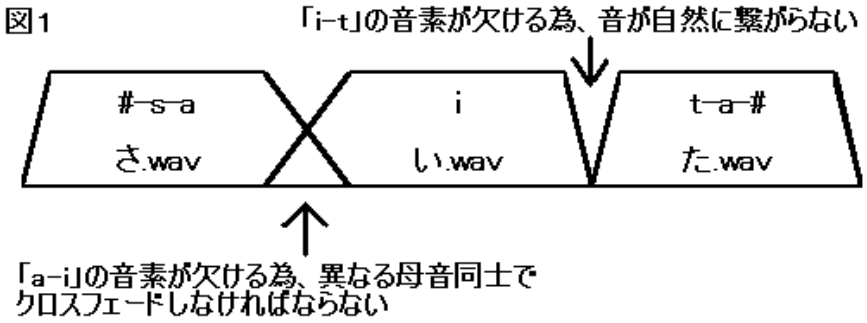

Figure 4: Consonant-Vowel diphone style recording. Notice the space between the last two syllables of the word.[11]

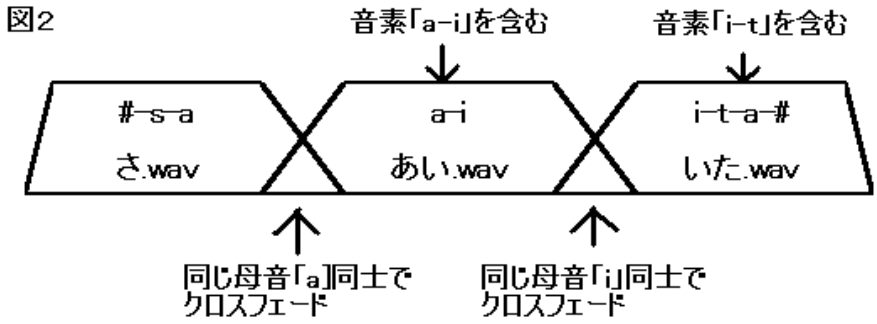

Figure 5: Vowel-Consonant-Vowel diphone style recording. The spaces between syllables is more easily crossfaded due to the previous vowel preceding the next sound.[11]

To add details, different parameters in the program can be added to a melody to modify it. These include but are not limited to: portamento, and pitch-bend, breathiness, and vibrato. These create a sense of realism and dimension for the vocals. In the Vocaloid community, this pitch articulation is often referred to as “tuning” a song. This process adds imperfections to the vocals to make them seem more human. The Vocaloid 4 engine also added a Pitch Rendering Feature to help visualize these portamentos.

The Vocaloid 4 engine also added the Growl and Cross Synthesis (XSY) parameters were added to increase this sense of realism. The growl function was a distortion parameter that was inspired by the "utaugrowl" resampler from the program UTAU, one of Vocaloid's many derivative programs. Yamaha advertised this feature with Vocaloid's fourth release, stating, "You can now apply a more human-like representation to your tunes in a wide variety of genres, whether it be intense rock shouts, hard core and death metals screams, or even soul music."[12] The XSY feature allowed for the blending of multiple voice libraries together during one song. This could be used to blend between a single character's different voice libraries, such as a natural and soft voicebank. This gave the program the capability to transition between multiple tones of voice during one song. These features were added to Vocaloid 5, the latest edition of the program, along with complex portamento generating presets to make the program more accessible to new users.

The Voice Behind the Curtain – Putting Faces to Names

Character design plays a big part in creating new Vocaloid products, as seen in the case of LEON and LOLA versus Miku, as previously mentioned. For many Vocaloid characters, their persona defines them more than their voice does. This goes to the extent of the fandom even defining the characteristics of their favorite characters by writing them a certain way in their music or depicting them in writing or art with other characters. Many Vocaloid 2 and 3 libraries were called "unoriginal" or "Miku copies" because their production companies aimed to make them marketable according to the fandom's standards. However, anything that doesn't fit this standard doesn't see massive success.

This often is related to identity politics surrounding the voice libraries' personas. This is best exemplified in talking about the original Vocaloid duo, LEON and LOLA. Even though Zero-G, the company that produced LEON and LOLA, made explicit efforts to have a voice for a black soul singer,[3] many of LOLA's users did not recognize LOLA as black, and these users posted complaints on Zero-G's message boards in response to it. One user even remarked, "Do we have a British soul singer with a Japanese accent who lisps like a Spaniard? Eesa makea me tho unhappy...."[3] To the credit of the user, it was a completely fair critique of the program's capabilities at the time, since LOLA's Vocals were incredibly choppy and hard to understand.

However, many users found her pronunciation to be strange, likely due to the Caribbean accent of the voice provider, as mentioned by one of the programmers. This generated many forum discussions about what genres the two should be used for, based on the inherent timbre of the voices, bringing up questions about the relationship between vocals and the body they belong to in a post-colonial world. Fifteen years later, the fanbase still hasn't gotten an official answer from Zero-G as to who voiced LEON and LOLA, though based on evidence from a demo video for the two voices and auditory comparisons, they are highly speculated to be Lance Ellington and Juliet Roberts,[13] both of whom are black British singers. A project manager for Zero-G has confirmed that both singers were black. Even with this evidence, LEON and LOLA often are whitewashed in popularized fan-generated depictions of the characters, more notably LEON, who is depicted as a white, blonde-haired and blue-eyed man.

Unfortunately, there are not many other dark-skinned or black coded Vocaloids, making this representation absolutely necessary. Out of about 90 Vocaloid characters,[6] LEON, LOLA, Merli, Wil (ZOLA Project), RUBY, AZAKA (Unity-Chan), Cyber Songman, Amy and Chris (Vocaloid 5 Standard Vocals) are the only dark-skinned or black Vocaloids, with the exception of RUBY, who is a lighter skinned Latina, as is her voice provider, Misha.

Wrapping Up

Vocaloid in its five versions fifteen years of existence has had a global impact on voice technology and has given a pathway for concatenative synthesis to come into the mainstream, with major entertainers featuring the phenomenon. The program is being used by fans and professionals alike which provides fans with user generated content in songs, animations, games, and more. Users strive to make their synthesized vocals sound more human with new parameters added to each new version of the program. However, because of the ability to give these anonymous voices a companion image, the image of the person behind each character's voice can be easily misconstrued or misrepresented.

Citations

1. Apol, 甘い Amαι. "All Vocaloid Release Dates and Box Art Part 1 | Vocaloid Amino." Vocaloid | Aminoapps.com. May 07, 2016. Accessed March 26, 2019. https://aminoapps.com/c/vocaloid/page/blog/all-vocaloid-release-dates-and-box-art-part-1/WJoU_XuewvwKbRmlozwRwxorZ1Loq2r. (Image)
2. "VOCALOID Development." Vocaloid Wiki. Accessed March 26, 2019. https://vocaloid.fandom.com/wiki/VOCALOID_development.
3. "Revista Transcultural De Música - Transcultural Music Review." TRANS. Accessed March 26, 2019. https://www.sibetrans.com/trans/articulo/57/synthesizing-race-towards-an-analy.
4. "Hatsune Miku." Wikipedia. March 15, 2019. Accessed March 26, 2019. https://en.wikipedia.org/wiki/Hatsune_Miku. (Image, voice provider name)
5. "Hatsune Miku Append." Vocaloid Wiki. Accessed March 26, 2019. https://vocaloid.fandom.com/wiki/Hatsune_Miku_Append.
6. "Status." Vocaloid Wiki. Accessed March 26, 2019. https://vocaloid.fandom.com/wiki/Status.
7. "Vocaloid." Wikipedia. March 17, 2019. Accessed March 26, 2019. https://en.wikipedia.org/wiki/Vocaloid#Technology.
8. Culture In Eastern Asia. "Japanese Language." Japanese Language. January 01, 1970. Accessed March 26, 2019. http://cultureineasternasia.blogspot.com/2014/10/japanese-language.html. (Hiragana Chart)
9. "IPA Pronunciation Key." Wiktionary. Accessed March 26, 2019. https://en.wiktionary.org/wiki/Wiktionary:IPA_pronunciation_key.
10. "Japanese Phonology." Wikipedia. March 02, 2019. Accessed March 26, 2019. https://en.wikipedia.org/wiki/Japanese_phonology.
11. "UTAU User Manual - 15." UTAU Wiki. Accessed March 26, 2019. https://utau.fandom.com/wiki/UTAU_User_Manual_-_15. (Diphone vs Triphone schematic figures)
12. Yamaha Corporation. "Try out These New VOCALOID4 Features! - What You Can Do with VOCALOID4." VOCALOID (ボーカロイド・ボカロ) 公式サイト. Accessed March 26, 2019. https://www.vocaloid.com/en/articles/vocaloid4_feature.
13. "LEON's & LOLA's Voice Providers Discovered?" UtaForum 4.0. Accessed March 26, 2019. http://utaforum.net/threads/leons-lolas-voice-providers-discovered.8324/.