Adam Billings

2 May 2021

The Efficacy of Music as a Constructed Language

Ever since the beginning of observable time, language has been humanity's way of communicating ideas, warnings, and complex emotions. Because of the constant and imperative use of language throughout history, a person can listen to a foreign speaker that they do not comprehend and still be aware that the speaker is communicating information that is meaningful. The reason for this is because of the intricacy of speech itself; the fluid organization of syllables, stresses of specific words, and the use of complex phonemes gives a human listener the idea that information is being communicated. These same basic aspects of language - rhythm, pitch, and timbre - are also the same building blocks on which music is created. While many composers directly use linguistic devices in vocal music, many elements of spoken language can be translated into extended musical techniques for specifically non-vocal instruments and ensembles. This can be done by analysing and comparing multiple linguistic rhythms, relationships between consonants and vowels, and pitched accents from many world languages in order to create a set of tools and techniques for communicative, non-vocal music composition.

One of the most important aspects of music since its inception has been the use of pitch. The formation of melody, harmony, and even rhythm are the result of the human ability to create and control different pitches. Many cultures around the world have varying degrees of pitch usage for communicating in their languages. Although pitch in many languages sounds different than a common musical melody, the techniques that are used can still be transcribed into a musical context. Some world languages are defined as *tonal* - meaning that the use of pitch can change the meaning of a single word - but in the English language the use of pitch creates a

context for the emotional weight of a sentence. Dwight Bolinger's essay, *A Theory of Pitch Accent in English* examines and identifies the use of pitch in the English language. Although Bolinger explains that English is not a *phonemic* language - meaning that a change of pitch during a word's pronunciation does not alter its meaning - he also points out that pitch does function similarly to the organization of vowels and consonants in English. For example, "a sequence of different pitches will produce something potentially meaningful in the same way that a sequence of vowels and consonants may produce a word" (Bolinger, 1958). It is shown that one can easily recognize the distinctive aural features between consonants and vowels, but the idea of pitch conveying meaning becomes impossible to discern, "unless it occurs in context" (Bolinger, 1958). Therefore, the role of pitch in the English language is in fact, to convey the context of a sentence.

Alternatively, tonal languages rely on pitch directly to discern different meanings of similar words. One single word can have multiple meanings based on the spoken melodic fragment that is used. Babbel - a popular language learning organization and magazine - explains that while English does use pitch for context, the Mandarin language uses pitch to change entire meanings of words. Thomas Devlin of Babbel uses the Mandarin word *ma* to show that "depending on the tone, *ma* can mean 'horse,' 'hemp,' 'scold' or 'mother.' If you want to say 'mother,' you'll say *ma* in a high tone, but if you say 'horse' you start high, then go down, then go up again in quick succession". Mandarin chinese uses four distinct pitch alterations and one pitch center in their spoken language (Devlin, 2019). This technique of tonal word modification is represented musically with the compositional idea of varied repetition. In a given piece of music, a melodic theme can be introduced purely in its most prime form. In Western music, this theme will then be used in multiple contexts in order to show development of the material.

Throughout the progression of the music, the theme will be recognized by the listener, but its use will be different each time. This differentiation between each iteration of the same material is similar to the use of pitch in tonal languages like Mandarin.

Going further, in Turkey, Kusdili is a Turkish-based language consisting completely of whistled melodic lines. In a video documentary by Great Big Story, it is explained that Kusdili is primarily used to give commands and warnings to inhabitants of Turkey's mountainous region of Kuşköy, meaning "Bird Village" (Gaynutdinova and Mutallimova, 2021; Great Big Story, 2020). Muazzez Köçek explains that there are different pitch ranges that can be used based on which fingers the whistler puts in their mouth. Explaining the primary use for the language, Köcek demonstrates a set of calls that translate to, "There is a wolf across the hill", "The fog of the mountain is coming down", and "Pick some apples from the tree". Often referred to as Turkish's Bird Language, Kusdili's melodic and rhythmic patterns sound remarkably similar to many standard bird calls. It is explained that the translation from Turkish to Kusdili is similar to the translation of Morse Code (Great Big Story, 2020). Similarly, in the Spanish island of La Gomera, inhabitants use a version of Spanish that is completely encoded into various whistled melodic lines. Like the Turkish Kusdili, Silbo Gomero's primary usage was "as a means of long-distance communication by most of the inhabitants of the mountainous interior to convey a wide variety of messages" (Rialland, 2003). Turkey's Kusdili and Spain's Silbo Gomero use melody to derive meaning in the same way that music does. Because of this, these languages can be translated directly into a musical context. The use of pitch in language has been used throughout the history of vocal music, where a sung melody will often be governed by the pitch contour of the spoken sentence. Mandarin's pitch alterations can be synthesized and transformed into the idea of musical motivic development, where one short melodic fragment can have

multiple musical contexts or meanings based on its transposition, pitch order, or its emphasized pitches. This does not require a human voice to draw a similarity between its musical and linguistic usages. Mandarin, along with Kusdili, Silbo Gomero and other tonal languages, can serve as melodic inspiration for instrumental music.

Another concrete feature of musical art is the use of rhythm. Rhythm has been used as a musical building block ever since the observable conception of music. Rhythm may not be directly inherent to music though, as almost all languages rely on rhythmic timing to communicate full sentences. A link between language and music can be found by analysing the relationship between the two. Linguistic rhythm is used in vocal music in order to create a sense of coherence in the musical setting of a particular text. That being said, the intricacies of rhythm in spoken language are still viable for instrumental transcription.

In *An Empirical Comparison of Rhythm in Language and Music*, Aniruddh Patel and Joseph Daniele approach the common suggestion that the way a specific culture's language is formatted has a large impact on the characteristics of the music from that culture. This suggestion is found by listening to the music of composers that were born in similar areas. Many academics have pointed out that there is a certain sound to French music, German music, Russian music and so on. Patel and Daniele aim to use mathematical evidence to find a rhythmic correlation between a composer's primary language and the music that they write.

They make the claim stating that it is difficult to analyse such a connection because of the abstraction of musical art. They state that the lack of empirical data "has been partly due to the difficulty of developing and applying comparable quantitative measures to melody and rhythm in speech and music" (Patel and Daniele, 2002). Patel and Daniele explain that there are two common ways to structurally analyse the connection between language and music: *prosodic*, and

syntactic. They explain that "comparisons of prosodic structure examine the way duration, pitch and intensity create structured rhythmic and melodic patterns in speech and music... [while] comparisons of syntactic structure examine the way discrete elements combine in hierarchical fashion to form coherent sequences in the two domains (Patel and Daniele, 2002). Likewise, in musical analysis, prosodic structure is similar to melodic structure, while syntactic structure is related to harmonic analysis. These two methods can be used to find commonalities between linguistic and musical analysis.

As a case study, Patel and Daniele compare the "stress-timed" linguistic rhythm of British English, with the "syllable-timed" rhythm of French. Patel and Daniele deduce a statement by Grabe & Low (2002) in saying that the evidence from comparing these two types of linguistic rhythms is that the duration of vowel pronunciation is *more* varied from word to word in stress-timed languages (like English) because these languages "show a greater degree of vowel reduction than syllable-timed languages" (Grabe & Low, 2002; Patel and Daniele, 2002). They go on to show how the monosyllabic word 'strengths' is an example of complex syllables that are widely used in stress-timed languages, coming to the claim that these languages (Patel and Daniele, 2002; Ramus et al., 1999). Because of this, Patele and Daniele explain Grabe and Low's deduction, further clarifying that "vocalic duration is more variable in British English, Dutch, and German (another stress-timed language) than in French and Spanish... [and] the temporal patterning of vowels is likely to be one factor contributing to the rhythmic feel of a language."

After discovering Grabe and Low's mathematical formula of finding the durational variability of a specific language (titled the "normalized Pairwise Variability Index, henceforth 'nPVI'), Patel and Daniele then apply this process to a highly specified sampling of music from

composers from England and France. In their selection, they were extremely diligent in filtering out music that could alter the results. These restrictions included removing all vocal music, all music with a specific "external rhythmic agenda (i.e. stylized dances such as waltzes, gigues, and gavottes, as well as marches)" (Patel and Daniele, 2002), all music that was composed outside of 1800-1900, and had melodic material that was at least 12 notes in length. The composers also had to be native speakers that were born in the country of their language (France or England). After analyzing 137 English and 181 French musical themes, they found that there was a notable difference between the rhythmic variability of French music to English music, with English music having a higher variability. These results corresponded to the results from analysing the nPVI of the two languages. The claim that the English language has a higher variability in linguistic rhythm than that of the French language holds true in the sampling of music from composers of their respective countries (Patel and Daniele, 2002). This result remains conclusive when applied to other stress-timed versus syllable-timed languages.

An Empirical Comparison provides many crucial details to further an understanding of the relationship between language and music in certain countries. By using real mathematical data to show the correlation between the rhythmic variability of melodic structures and the mother languages of the composers that wrote them, one can use these findings in a new musical application. This technique is used inherently in almost all vocal works, where the stresses of different words dictate their placement in a musical meter. Though, one could go further and use this technique in a more subjective manner with non-vocal instrumental ensembles. By understanding the differences in the flows of stress-timed and syllable-timed languages, a composer could, for example, create contrasting musical character pieces that utilize the rhythmic contrasts between their representative languages. The following is a proposed example



of an attempt to notate the rhythmic difference between the same sentence in English and French.

These two rhythms could be used as motives in a larger piece contrasting the two linguistic rhythms.

Another language built entirely on the idea of rhythmic communication is Morse Code. Originally invented by Samuel F.B. Morse in the United States in the 1830s, Morse Code was used primarily for electrical telegraphy (Encyclopædia Britannica). The language originally encoded English into a series of written dots and dashes. This binary relationship was auditorily transcribed into a short-long system. Musical rhythm can be derived from translating Morse Code directly. For example, this is a transcription in common meter using eighth notes as the short rhythm, quarter notes as the long rhythm, and eighth rests to signify the start of a new letter (This transcription system was derived from David and Phil Tulga's *Morse Code Music*):



In a musical context, the difference between the sound quality of different instruments is referred to as *timbre*. If a clarinet and a flute play the same note, the differentiation between the sound that identifies one as clarinet and the other as flute is the *timbre* of the instrument. One

singular instrument can have multiple timbres, as well. This is done by altering the way one plays the instrument to produce different qualities of their original tone. The same can be said about spoken language. Khoisan, an African language family consists of many tongue clicks and tongue pops to pronounce vowels. These clicks and pops add another level of articulation and accent to many words in the Khoisan linguistic family. Some Khoisan languages use the latin alphabet in writing, and because of this, many words look similar when written (Khoisan Language Family). A majority of Khoisan languages use five distinct clicks, phonetically titled *bilabial, dental, alveolar, lateral,* and *palatal*. These clicks are onomatopoetically described as follows:

"A *bilabial* click is a pop made by bringing the lips together and releasing them, just like the sound of a kiss... A *dental* click sounds like 'tsk, tsk! ' and is made by putting the tongue just behind the front teeth... An *alveolar* click sounds like the popping of a cork made by putting the tongue just behind the ridge in the back of the front teeth... A *lateral* click sounds like the sound used in English to urge on a horse... [and] A *palatal* click is a sharp pop made by drawing the tongue down quickly from the roof of the mouth." (Khoisan Language Family)

From initially listening to the Khoisan languages, the popping/clicking accents are quite surprising and offer a new and fresh timbre to the traditional linguistic properties that are ubiquitous in most other languages. The seemingly consistent uses of these consonant sounds add to the idea that even without fully understanding the meaning of the words, it is still obvious that there is meaning being communicated. The essence of the consonant timbral accents found in the Khoisan languages can be translated into a musical context. Many well-known woodwind extended techniques allow the performer to create similarly accented popping sounds out of their

instrument. For example, the sound of a flute 'tongue-ram' is often referenced as sounding like the popping of a cork, exactly like the description of the *alveolar click* (Khoisan Language Family). The quality of many of the special Khoisan consonants are also similar to the accented sounds of percussion instruments. The *palatal* click possesses a timbre that might be reproduced by striking an orchestral wood block. Likewise, the dental click could be musically represented by a quick swipe of a güiro. This specific translation of language into musical context provides a use for this accented and timbral variety.

In order to find particular congruences between languages, one must know how they might be created. Many movies, books, and video games often have languages that do not exist outside of the fictional world. These constructed languages (often referred to as *conlangs*) are a product of the search for an all-encompassing set of linguistic devices that separate communication from nonsense. In the essay, Ludic Linguistics: A Revisited Taxonomy of Fictional Constructed Language Design Approach for Video Games, Luthfie Arguby Purnomo (et al.) dissects constructed languages (conlangs) into two categories: "priori, fictional conlang design not based on any existing languages, a *posteriori*, conlang design adapted from existing languages" (Purnomo, et al. 2017). In a study comparing conlangs to the real languages of their creators, Sara Blalock Ng, and Abigail Schwendiman concluded that many constructed languages "rely on the relationship between a conlang and the native language(s) of its author. On average, 62.42% of the segments in the conlang inventories were also present in their parent inventories" (Ng & Schwendiman, 2017). The profession and documentation of "conlanging" becomes useful when attempting to create a form of communication through music. At The Massachusetts Institute of Technology, students enrolled in the course: ConLangs: How to Construct a Language, created languages that use an extremely limited set of sound sources for

Billings 10

communication. For example, "one student... created a language for underwater creatures who speak in shades of color" (Humanities, 2019). Because of the successful use of these limitations in building an effective form of communication, it becomes clear that the elements of musical rhythm, pitch and timbre can also be utilized to create an intricate constructed language.

In Text-to-Speech Synthesis, Paul Taylor explains the fundamentals of communication. This is quite important when attempting to create communicative music. Taylor explains affective communication: "where we express a primary emotional state with external means. A good example of this is the expression of pain, where we might let out a yell or cry upon hurting ourselves." He explains that the most important characteristic of affective communication is that the "intensity of the external form is clearly a function of the intensity of feeling; the more intense the pain the louder the yell" (Taylor, 2009). The most primary emotions like happiness, anger and sadness are commonly expressed using affective communication. This type of communication is a primary compositional technique in music. In elementary school, students learn the difference between major and minor chords based on the surface-level affective emotions the chords possess. A common trope in primary music education is that the major chord resembles happiness, while the minor chord resembles sadness. It can be inferred that nearly all composers in the history of Western music have built on these fundamentals, using the idea that "the more intense the pain, the louder the yell" (Taylor, 2009) as a validation for dynamic contrast in many musical scenarios. In a video from WIRED, musician Jacob Collier takes this idea to a musical extreme, by using the famous tune Londonderry Air (also known as Danny Boy) and transforming it in order to musically depict eighteen increasingly abstract emotional states. Beginning with the simplest of emotions, *happy* and *sad*, and traveling in stages through more subjective emotions like guilt, flirtatious, and inevitability (WIRED, 2020),

Jacob's modern "theme and variations" on Londonderry Air is an elevated example of affective communication in music.

Alternatively, symbolic communication "give[s] up the idea that the form must indicate the meaning. Rather we use a series of correspondences between form and meaning, in which the relationship is not direct" (Taylor, 2009). Taylor gives an example of someone communicating that the act of showing his left hand means 'tiger', while showing his right hand means 'lion'. This communicates to the listener a visual difference between the two symbols of tiger and lion. Taylor argues that "the disadvantage is that the fact that left-arm-wave means tiger, whereas right-arm-wave means lion, has to be learned" (Taylor, 2009). This is exemplified in music, using the compositional idea of a motive. The term, leitmotif - famously coined by Richard Wagner - is a "theme of easily recognizable melodic, rhythmic or harmonic identity, first used in connection with a certain character of incident, and which returns time and again, always with a reminiscence of the original association" (McShan, 1997). No matter what the exact musical content of the leitmotif is, it is inevitably arbitrary. It is only up to the composer to use techniques like varied repetition and development to make the audience recognize the phrase as a memorable and important one. Many composers have used this technique, as it is almost essential to the assured coherence of any time-based subjective art form. This idea of leitmotif is commonly used in the film music industry, where composers use thematic material to relate to the character development in the movie. Some specific film scoring techniques have been so ubiquitously used that they often become cliché. When this happens, even untrained audiences can recognize the commonalities of the current theme and relate it to themes that have been used in similar contexts. This relates back to the most primitive associations of "major means happy," and minor means sad" that have been an important part of western primary music education.

These common tropes can be extrapolated into a foundation for a form of musical communication that inherently resonates with many first-time-listeners.

In *The perception of rhythm in language*, Anne Cutler points out the separation of written and spoken language. Most written languages have spaces between every word, but these spaces are not communicated when speaking the language. The listener does not expect a pause between each word in order to clearly understand the meaning of the sentence (Cutler, 1994). This is related to Bolinger's remarks in *A Theory of Pitch Accent in English*, showing that the difference in perceived information between text and speech comes from the addition of a speaker's emphasis on certain syllables, words, and even sentences. The contextual clues that are evoked in speech gives a heightened understanding of the information being communicated - something that can't happen without auration. This same problem happens with composed music, where the written notation shows the audience all of the information that they will hear, but gives only an inconclusive idea of what the audible context of the piece will possess. Only the most trained musicians and musicologists can fully realize a written score in their head, but even then the entire aspect of performance is still lost. The lack of standardized post-modern musical notation often creates miscommunications between the composer and performer as well.

One of the main distinctions between music and language is how they are learned. With language, in order to associate a meaning to a spoken sound, one must learn the meaning first, while with music, one doesn't inherently *need* to learn anything beforehand. In an interview on *60 Minutes*, Paul McCartney of The Beatles explains: "I don't read music or write music. None of us did in the Beatles. We did some good stuff though. But none of it was written down by us. It's basically notation. That's the bit I can't do. 'Cause I don't see music like that. I don't'' (Alfonsi, 2019). It becomes difficult to create a concrete system for musical communication,

simply because of the idea that "a word is in effect a form/meaning pair in which the correspondences between the meaning and form is arbitrary. Because of this, there is no way to guess the meaning of a word if it is not learned" (Taylor, 2009). Every communicative word must be learned first before understanding what the word means. The interest comes from listening to languages that one does not know. For example, if a person only knows the English language, but they are listening to French, they may not understand any of the words, but they will understand that the French speaker is communicating something of meaning. The fluid organization of consonants and vowels, the rhythm of each sentence (pauses, cadences), and the accents and pitch shifts on specific words exemplifies the intricacy of the spoken language and informs even non-native speakers that there is detailed communication that is happening. Using the previously mentioned connections between linguistic and musical techniques in combination with affective communication (expression of a primary emotional state with external means) can transform this phenomenon of recognizing foreign communication into a musical form.

Although it may be difficult to associate complex definitions or meanings to certain musical motives, the idea of creating a system of communication through music (similar to a constructed language) is possible if the composer defines the meanings and associated musical identifiers to the audience before a performance. This can be done in multiple ways. For example, the composer could give a presentation or table defining each musical word or sentence they have created, and how it will function in their piece. In order to create a standalone work, though, the composer would need to establish the definitions through repetition, and possibly even include visual identifiers that the performer can use to further convince the audience of the intended meaning behind the musical gesture. This is a form of symbolic communication, as defined by Paul Talor in *Text-to-Speech Synthesis*.

Ultimately, the methods of creating a loose musical version of communication versus creating a fully defined musical constructed language are two fully validated approaches for achieving musical communication. Composers can use these aforementioned techniques to form a new style of instrumental music. The creation of a musical language would require composers and musicians to collaborate on an undisputed set of linguistic properties and their associated musical meanings. Doing this would open the door for many new ways to utilize melody, rhythm and timbre, allowing for implementations of non-vocal recitative-style works, meaningfully associated character pieces, as well as a fully formed in-depth musical language. The foundation of affective communication that composers have built for hundreds of years can now be expanded on. An intricate and communicative musical language is the natural progression of the already existing emotional connection between the composer, the performer, and the audience.

Bibliography

Alfonsi, Sharyn. "Paul McCartney Opens up about Abbey Road, the Beatles' Breakup in Wide-Ranging Interview." CBS News, CBS Interactive,

Bolinger, Dwight L. "A theory of pitch accent in English." Word14.2-3 (1958): 109-149.

Cutler, Anne. "The perception of rhythm in language." (1994).

Devlin, Thomas Moore. "What Are Tonal Languages?" *Babbel Magazine*, 28 August, 2019 www.babbel.com/en/magazine/tonal-languages.

"Morse Code." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., www.britannica.com/topic/Morse-Code.

- Gaynutdinova, Alina S.; Aliya Mutallimova. "The Culture and Language of Whistle of Turkish People (Giresun)". *International Journal of Society, Culture & Language*, 9 January, 2021, 86-93.
- Great Big Story. "This Turkish Language Isn't Spoken, It's Whistled." *YouTube*, YouTube, 21 Jan. 2020, www.youtube.com/watch?v=l117wfB0g3o.
- Humanities, School of. "How to Build a Language." *MIT News* | *Massachusetts Institute of Technology*, news.mit.edu/2019/constructed-languages-linguistics-1218.

"Khoisan Language Family - Structure &; Writing - MustGo." *MustGo.com*, www.mustgo.com/worldlanguages/khoisan-language-family/.

McShan, Jessica K. *Wagner and the Leitmotiv*, umich.edu/~umfandsf/symbolismproject/ symbolism.html/Teutonic_Mythology/wagleit.html. 17 December, 1997.

www.cbsnews.com/news/paul-mccartney-opens-up-about-abbey-road-the-beatles-breaku p-yoko-ono-in-wide-ranging-60-minutes-interview-2019-06-23/. 23 June 2019.

- Ng, Sara Blalock, and Abigail Schwendiman. "Properties of Constructed Language Phonological Inventories." *University of Washington* 2 (2017).
- Patel, Aniruddh D., and Joseph R. Daniele. "An empirical comparison of rhythm in language and music." *Cognition* 87.1 (2003): B35-B45.
- Purnomo, SF Luthfie Arguby, et al. "Ludic Linguistics: A Revisited Taxonomy of Fictional Constructed Language Design Approach for Video Games." *GEMA Online Journal of Language Studies*, ejournals.ukm.my/gema/article/view/18749.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. Cognition, 73, 265–292.
- Rialland, Annie. "A new perspective on Silbo Gomero." Proc. 15th ICPhS. 2003.
- Taylor, Paul. Text-to-speech synthesis. Cambridge university press, 2009.
- Tulga, David and Phil. Morse Code Music. www.philtulga.com/morse.html. 2004-2021
- WIRED. "Jacob Collier Plays the Same Song In 18 Increasingly Complex Emotions | WIRED." *YouTube*, YouTube, 6 Oct. 2020, www.youtube.com/watch?v=EWHpdmDHrn8.