Vocal Imitation and Linguistic Processing of Super Mario Theme Music by Yorùbá Gamers

Samuel Akinbo
University of British Columbia

Abstract

An aspect of gaming culture among Yorùbá millennials is the linguistic interpretations of the background music that accompanies the popular video game called Super Mario. The themes of the linguistic interpretations are comparable to those of music texts at traditional Yorùbá competitions (Apter, 1998; Fámulo, 2018). Drawing on the Yorùbá music tradition, the account in this work is that the gamers assumed that the background music of the video game performs a similar function as the music at traditional Yorùbá competitions. Semantically, the choice of words in the linguistic interpretation is conditioned by the situational contexts or scenes where the music is heard in the video game. The results of an acoustic analysis show that the choice of words in the linguistic interpretations are also determined by mapping the pitch trajectories of the music melodies to the tones of the gamers’ native language. Notably, the results of this study suggest that the linguistic processing of music may not only involve phonetic iconicity (Hacohen and Wagner, 1997; Koelsch et al., 2004; Steinbeis and Koelsch, 2011) but situational context and social expectation (Feld, 1984; Agawu, 2001; Patel, 2008; García et al., 2014; Neumeyer, 2015).

1. Introduction

Ancient philosophers and thinkers have contemplated the relations between language and music. For example, the Greek philosopher Plato claimed that the power of music stems from its resemblance to the sounds of speech (Neubauer 1986, cited in Patel 2008). Most strikingly, the music-language connection is embodied in the capability of the talking drums or flutes in many African music traditions. The most popular example of talking drum comes from Yorùbá which is a Benue-Congo language with more than 20 million speakers in South-Western Nigeria (Pulleyblank, 2009). Like most Nigerian languages, Yorùbá is a tone language, which means pitch contrasts bring about lexical or grammatical distinctions in meaning (Yip, 2002). The language has three contrastive tones: H(igh) bá ‘meet’, L(ow) b `a ‘land’, and M(id) ba ‘to plait’ (Bamgbọse, 1967; Awóbùlúyì, 1978; Akinlabi, 1985; Pulleyblank, 1986). Yorùbá musicians communicate through a family of Yorùbá drums called dundún by varying the compression of the drum’s tension cords. The varying compressions of the tension cords encodes the pitch contours of tones in Yorùbá words (Beier, 1954; Euba, 1990; Waterman, 2000; Villepastour, 2010, 2014). Using acoustic analysis, Akinbo (2019) shows that there is a positive correlation between the pitch contours of the speech tones and those of their corresponding musical rendition (Figure 1).
Figure 1: Mean $F_0$ contours for Yorùbá drum and speech tones

Figure 1 shows the drummed and spoken forms of the Yorùbá phrase [ògúnbòdè gbàgbá] “Ògúnbòdè took garden egg”. In the graph, the y-axis contains the acoustic measurement of pitch contour in $F_0$(Hz), and the x-axis contains the tones of the Yorùbá phrase. The dark line is for the drum pitch and the grey line is for the speech pitch. As shown in the graph, the pitch contours of the Yorùbá phrase and the corresponding drum rendition have similar trajectory.

Just as in Yorùbá, the musicians in other cultures around the world are also able to musically communicate by encoding meaning or the phonemic aspects of their languages with music instruments (Stern, 1957; Carrington, 1971; Ekwueme, 1974; Bradley, 1979; Lo-Bamijoko, 1987; Poss, 2005; Winter, 2014; Poss, 2012; Seifart et al., 2018; McPherson, 2018). Studies on music communication in cultures with musical speech surrogates have tended to focus more on musical imitation of language, but it is also important to examine the “opposite”, which involves language imitation or interpretation of music melodies.

Musicologists and cognitive psychologists have long been investigating how musical melodies are linguistically interpreted (Feld, 1984; Hacohen and Wagner, 1997; Koelsch et al., 2004; Patel, 2008; García et al., 2014; Neumeyer, 2015), but most of these studies focused on music traditions without speech surrogate systems. In a culture with musical speech surrogates, studying musical meaning from a linguistic perspective could offer valuable insights into the role of the acoustic parallel between language and instrumental music in music cognition, and the connection between musical and linguistic communication. Using linguistic instrumentation and methodology, the present study contributes to research on musical meaning. Specifically, this study explores how Yorùbá speakers linguistically interpret unfamiliar instrumental music.

The present work is based on conventionalised interpretations of the background music in the Super Mario game by Yorùbá gamers. The linguistic interpretations are discussed in §2. The discussion in §3 explores why the gamers interpreted the video game music
into Yorùbá. To compare the acoustics of the linguistic interpretation to the acoustics of
the corresponding background music, an acoustic investigation was conducted. Details of
the acoustic investigation are presented in §4. The results of the acoustic investigation are
presented in §5. The role of situational contexts in the linguistic interpretation is explored
in §6. In §7, the interpretative moves are analysed within the pragmatic theory of musical
meaning (Feld, 1984; Patel, 2008). The summary and the conclusion are presented in §8.

2. Interpreting the theme music of Super Mario Bros.

The data source for the present study is the linguistic interpretation of Super Mario music.
In this section, I present background information about the game, the theme music of the
game and the linguistic interpretation of the theme music by Yorùbá gamers.

Nintendo, which is a multinational Japanese company, released the console video
game called Super Mario Bros in 1985. More than 40 million copies of the game were sold
worldwide (Stuart, 2010). The protagonist of the game is Mario who is an Italian plumber.
In the game which is set in the Mushroom Kingdom, the player takes on the role of Mario
or his brother Luigi in a multiplayer mode. The objective of the game is for Mario (or
Luigi) to save Princess Toadstool, but for Mario to save the princess, he needs to survive
the main antagonist Bowser, the forces of Bowser, the dangerous terrains in the game and
the allotted game time (Nintendo, 1985; Loguidice and Barton, 2012).

The pianist, Koji Kondo, composed the theme music of the game. According to
Laroche (2012), the composition of the theme music is a musical reaction to “players’
tended impressions”. Following Collins (2009: 6), the sounds in Super Mario Bros can
be classified into interactive and adaptive audios. Interactive audios are “sound events
directly triggered by the player’s input device”. The sounds which accompany jumping,
hitting an object, shooting, etc. are examples of interactive audios (see Lerner, 2014). On
the other hand, adaptive audios are not controlled by the player. Adaptive audio events
are cued by the game scenes, locations, game time, the presence of non-player characters,
etc. There are at least eighteen sounds in the game, but the focus of the present paper is
four adaptive audio events, namely Overworld/main theme, Flagpole Fanfare, Underworld
theme and Death theme. Because the Overworld theme is played throughout the game with
some tempo variations in certain stages, it is also called the main theme (see Lerner, 2014;
Schartmann, 2015). The musical experience in video games are the major innovations of
Super Mario Bros (Collins, 2009; Lerner, 2014).

It is uncertain when the game got to Nigeria, but the game was very popular in Nigeria
around 1990s and early 2000s. An aspect of gaming culture among Yorùbá gamers in Nige-
ria is the linguistic interpretation of Super Mario instrumentals into Yorùbá (Ayoola, 2019).
The linguistic interpretations developed naturally in the gaming community. Among all the
instrumentals in the game, only the adaptive sounds were conventionalised by the gamers.
This paper only focuses on the conventionalised interpretations. The conventionalised in-
terpretation of the Overworld theme is presented in Figure 2.
The Overworld theme, played at the first stage of the game, is also called the main theme. The linguistic interpretation of this theme is an adaptation of a Yoruba idiom used as a death threat. However, in this interpretation, the phrase /gün 'yán/ 'pound iyán' in the idiom is replaced by /tè 'bà/ 'make èbà' (lit: mash èbà').

When Mario is about to enter the castle after winning the first stage of the game, the Flagpole fanfare is played. The linguistic interpretation of this theme by Yoruba gamers is presented in Figure 3.

The second stage of Super Mario Bros is set in the underworld. Immediately upon entering the underworld, the Underworld theme is played. The first six notes of the theme are repeated once, but the notes and their repetition have different linguistic interpretations. The linguistic interpretations of these rhythms are presented in Figure 4.

---

1Iyán is the Yoruba name for pounded yam; èbà is a staple food made from dried grated cassava (manioc).
If Mario dies at any stage of the game, regardless of the cause, the Death theme is played. The gamers also linguistically interpret this theme music and the linguistic interpretation is presented in Figure 5.

In general, Yorùbá listeners (i.e Super Mario gamers) are able to linguistically translate or interpret certain portions of some theme music in Super Mario Bros. The interpretations are usually performed as a cappella. Every Yorùbá gamer who has played Super Mario knows these conventionalised interpretations, but it is uncertain whether someone who has not heard the theme music before might have the same interpretations.

An interesting feature of the linguistic interpretations is that a lexical syllable is mapped to each music note of the theme music. However, the choice of tone on a syllable is unpredictable. So, it will be interesting to investigate the factors that determine the choice of tone on the syllables. Although studies on Yorùbá speech surrogates show that the pitch contours of the lexical tones with their phonetic realisations are encoded with Yorùbá drums (Villepastour, 2010; Akinbo, 2019), it remains to be seen if these features are used in the interpretation of the theme music in Super Mario Bros. These issues are addressed in §4. Before focusing on the tone of the syllables, the next section addresses why the gamers assume that the background music of the game are communicative.

3. Why linguistically interpret the theme music?

The discussion so far has shown that the gamers linguistically interpreted the theme music of Super Mario by mapping lexical syllables to the music notes. Instead of mapping lexical syllables to the music melodies of the theme music in Super Mario Bros, the Yorùbá gamers
could have mapped non-lexical vocables or nonsense syllables to the music themes, as is
done in other cultures (see Hughes, 2000; Mullins, 2014; Weir, 2015). So, why did the
gamers use lexical syllables instead of vocables? In other words, why do the players assume
that an unfamiliar music involves linguistic communication? To answer this question, we
must understand the role of music in Yorùbá traditional games and events.

In Yorùbá culture, as in other African cultures, music is functional, contains textual
components, and plays an important role in traditional games and events (Adedeji, 1972;
Apter, 1998; Agawu, 2001; Green, 2005; Omojola, 2011; Campbell, 2015; Agawu, 2016).
For instance, in Yorùbá societies, instrumental and vocal music often feature in wrestling
matches such as those of Òròyéyé festival in Òyéjé, Èkitì state, Nigeria. The background
music is played by the supporters from the opposing sides. The background music either
(1a) threatens, (1b) warns or (1c) mocks the wrestlers. The general goal of the music is to
deter the opponent from being victorious. As Apter (1998) rightly puts it, the wrestlers fight
with their bodies, but their supporters fight with words. Similar behaviours are found in
non-African events and games (see Lee, 1985; McLeod, 2006; Vale and Fernandes, 2018).

(1) Songs from Òròyéyé festival Wrestling matches (source Apter, 1998)

<table>
<thead>
<tr>
<th>Type</th>
<th>Song</th>
</tr>
</thead>
</table>
| Threat  | bírí gbe. ‘be carried by the wind’
        | òyì gbe... ‘be dizzy’                                               |
| Warning | omodé é gun’ gi ọgèdè, ‘a child that climbs the banana tree,’
        | a yó báárá, ase wì... ‘will slide down’                            |
| Mockery | Adé ró di, ‘the crown has thumped down,’
        | bí ọní koyín ‘like a bunch of palm kernels’                        |

Musical instruments such as the “talking drum” also play similar roles as vocal music
at traditional Yorùbá events. For example, during masquerade festivals, Yorùbá drummers
often play provocative phrases on the talking drum in order to excite the masquerades or
the masqueraders. When the phrases in (2) are played on the drum (or spoken by the spec-
tators), the masqueraders intensify their dance or any other action. In this sense, vocal and
instrumental music at traditional Yorùbá events are discourse (Agawu, 2001; Villepastour,

(2) Yorùbá drum phrases (Fâmúlé, 2018)

<table>
<thead>
<tr>
<th>Type</th>
<th>Phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>oò le sê</td>
<td>ô bá bá a rẹ ń sẹ ‘you are not as competent as your father’</td>
</tr>
<tr>
<td>b’óbá sê</td>
<td>‘If I were you’</td>
</tr>
<tr>
<td>ñ bá fápá</td>
<td>‘I will dance with my hands unceasingly’</td>
</tr>
</tbody>
</table>

Thematically, the linguistic interpretations of the video game music are comparable
to those of the background music in traditional Yorùbá games and entertainments. For
the Yorùbá gamers, the Overworld and Underworld themes are death threats to the player-
controlled character or more specifically the player. While the Death theme mocks the
death of the player-controlled character, the Flagpole theme is a celebratory theme for the
victory of the player-controlled character. Based on the thematic similarities, we can say that the gamers consumed the video game music using their background knowledge about the functions of background music in Yorùbá recreational activities. In other words, the interpretive move of the gamers can be considered an effect of social expectation, which is “an internalised social norm for individuals and organizations...about what people should do” (Hasegawa et al. 2007: 180).

That the Yoruba gamers utilised their cultural knowledge about game music in their interpretation of the video game music is in line with the enculturation account of musical interpretation, which suggests that musical meaning is exclusively determined by cultural convention and social background (Keil and Feld, 1994; Walker, 1996; Gregory and Varney, 1996; Balkwill and Thompson, 1999). For example, Feld (1984) suggests that a listener might relate a musical object or event to personal and social conditions, and related experiences where a similar sound object can be heard. However, Gregory and Varney (1996) suggests that the interpretation of music is determined more by cultural tradition than by the inherent qualities of the music. Social expectations have also been shown to facilitate comprehension and evaluation of spoken language (Rubin, 1992; Devos and Banaji, 2005; Kang and Rubin, 2009; Yi et al., 2013; Babel and Russell, 2015; McGowan, 2015).

In addition to social expectation, the physical properties of music such as pitch and tempo have also been shown to prime the interpretation of music. For example, faster tempo are associated with joy and happiness for adult listeners (Scherer and Oshinsky, 1977; Kellaris and Kent, 1993; Gagnon and Peretz, 2003; Webster and Weir, 2005; Eerola et al., 2013). Similar results are found for children music listeners (Dalla Bella et al., 2001; Mote, 2011). However, these studies are based on music traditions without speech surrogates. Considering that Yorùbá has a music tradition with a speech surrogate system which mostly relies on mapping the pitch contours of speech tones to music melodies (Villepastour, 2010, 2014; Akinbo, 2019), it is important to inquire whether pitch contours of music melodies influence the interpretation of music. Specifically, does the music interpretation involve vocal imitation of non-linguistic sounds? For this purpose, we must compare the pitch contours of the music notes in Super Mario to those of their conventionalised linguistic interpretation in Yorùbá. This investigation is crucial to tones of the lexical syllables that are mapped to each music note. In order to compare the pitch contours of the theme music and the linguistic interpretation, I conducted an acoustic investigation on the pitch contours of the conventionalised interpretations and the theme music.

Before turning to the acoustic investigation, the summary of the discussion here is that Yorùbá gamers assume that the background music of the game is communicative as a result of drawing on the roles of background music at traditional Yorùbá games. In this case, the cultural background and experience of the gamers are motivations for the linguistic interpretation of the instrumental music.
4. Methodology

Ten native speakers of Yorùbá were recruited in Vancouver, British Columbia for this study. The participants are 4 females and 6 males who immigrated from Nigeria and have spent at most seven years in Canada. They were all between the ages of 26 and 45. The stimuli in this study are the relevant Super Mario music and their conventionalised linguistic interpretations in Yorùbá. Because the sounds in Super Mario Bros often overlap, the relevant Super Mario theme music were extracted from YouTube videos of the instrumental music (sources Youcanplayit, 2012a,b,c; Luuul’s, 2016). Just as in the game, the themes were performed on a piano by the YouTube sources.

The linguistic interpretations with and without the corresponding instrumental music of Super Mario were presented to each of the participants. When the participants were presented the linguistic interpretations without the instrumental music, they were instructed to read the linguistic interpretations in speech mode. When the linguistic interpretations were presented with the instrumental music, the participants were instructed to produce the linguistic interpretations in a cappella mode. Each stimulus was repeated three times in speech mode as well as in a cappella mode. The data were recorded with a SHURE WH30XLR cardioid condenser (a headset microphone) at the sampling rate of 48.1 kHz in .wav format. The microphone was attached to a zoomQ8 camera. The vowels in speech and a cappella renditions were annotated in Praat (Boersma, 2001). The music notes of the theme music were also annotated in Praat. F0 values of the annotated intervals were automatically extracted at 50% point for the Super Mario Bros themes and their linguistic interpretations. In the next section, I present the results of the production experiment.

5. Pitch of Super Mario music and the corresponding Yorùbá interpretations

The F0 contours of the Super Mario theme music and their corresponding linguistic interpretation are presented in this section. We will start the discussion by looking at the Underworld theme. As shown earlier, the first-six notes of Underworld theme are repeated once (see §2). The pitch contour of the Underworld tune and that of its repetition are compared with their corresponding linguistic interpretations. For ease of identification, the initial portion of the Underworld theme is labelled Underworld1 and its repetition is labeled Underworld2. The line plot of Underworld1 music notes and the a cappella of its linguistic interpretation are presented in Figure 6. The y-axis contains the F0 values, and the x-axis contains the string of speech tones or music notes. The results from each participant are grouped by panels (i.e. p1, p3, p#).
Figure 6: Pitch contours of Underworld1 in Mario and a cappella mode

The F0 contours of Underworld1 notes and those of the linguistic interpretation in an a cappella rendition are plotted in Figure 6. Apart from individual differences among the participants, the F0 contours of Underworld1 is similar to those of their linguistic interpretation in an a cappella mode for all the participants.

To understand the strength of the relationship between the F0 of Underworld1 notes and the corresponding linguistic interpretation in an a cappella mode, a correlation coefficient of the speech and music tones were calculated using ggpubr (Kassambara, 2018). For this purpose, the mean of F0 values for all the participants was plotted against F0 values of the music tones. The y-axis contains the F0 values for the Underworld1 notes and the x-axis contains the mean F0 values of the corresponding tones in a cappella mode. The mean values of Underworld1 notes for all the participants in the game are plotted against those of their corresponding a cappella mode. This is repeated for the other theme music and their corresponding interpretations. The result of the Underworld1 theme is presented in Figure 7. For the reasons of formatting and for having similar results, the correlation coefficient for Mario instrumentals and their interpretations in speech mode are only presented at the end of this section.
Figure 7: Mean $F_0$ correlation between Underworld1 of Mario and a cappella mode

Figure 7 shows that there is a high degree of positive relationship ($R=0.97$) between the mean $F_0$ of Underworld1 notes and that of their corresponding linguistic interpretations in a cappella mode, and that this relationship is significant ($p=0.0013$). I now turn to $F_0$ contours of Underworld2, which is based on the repetition of the music notes in Underworld1.

Figure 8: Pitch contours of Underworld2 in Mario and a cappella mode

Figure 8 shows that the $F_0$ contours of Underworld2 notes and those of their linguistic interpretation in a cappella mode are similar. The degree of relationship between the $F_0$ values of the Underworld2 notes and those of the tone in their corresponding a cappella mode are presented in Figure 9.
Figure 9: Mean $F_0$ correlation between Underworld2 of Mario and a cappella mode

As shown in Figure 9, there is a high degree of positive relationship ($R=0.95, p=0.0031$) between the Underworld2 notes and their interpretation in a cappella mode.

Figure 10: Pitch contours of Overworld in Mario and a cappella mode

As shown in Figure 10, the trajectory of mean $F_0$ contours for the Overworld theme is similar to those of the linguistic interpretation in a cappella mode.
Figure 11: Mean F₀ correlation between Overworld of Mario and a cappella mode

Figure 11 shows that there is a high degree of positive relationship ($R=0.79$) between the mean F₀ of Overworld notes and the corresponding linguistic interpretation in an a cappella mode, and that this relationship is statistically significant ($p=0.0062$).

Figure 12: Pitch contours of Death theme in Mario and a cappella mode

Just as the aforementioned cases, the mean F₀ contour of the Death theme is similar to those of the linguistic interpretation in an a cappella mode.
As shown in Figure 13, there is a high degree of positive relationship ($R=0.72$) between the mean $F_0$ of Death theme and the corresponding linguistic interpretations in an a cappella mode, but the relationship is not statistically significant ($p=0.069$).

Unlike the results of the previous theme music, the mean $F_0$ contour of Flagpole theme and that of the linguistic interpretation are mainly similar at the beginning and the end. The strength of the relationship between mean $F_0$ of Flagpole theme and that of the corresponding linguistic interpretation in a cappella mode is shown in Figure 15.
As shown in Figure 15, there is a low degree of positive relationship between mean F₀ values of the Flagpole notes and those of the corresponding linguistic interpretation in a cappella mode (p=0.17, p=0.069).

So far, we have only compared Mario theme music to a cappella mode of the corresponding linguistic interpretations. This is because the results are similar when we compare Mario theme music to the speech of the corresponding linguistic interpretations. The summary of all the results are presented in Table 1.

### Table 1: Correlation coefficients of the music notes and their linguistic interpretations

<table>
<thead>
<tr>
<th></th>
<th>S vs A</th>
<th>M vs S</th>
<th>M vs A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underworld1</td>
<td>R 0.97</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>p 0.0011</td>
<td>0.0046</td>
<td>0.0013</td>
</tr>
<tr>
<td>Underworld2</td>
<td>R 0.95</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>p 0.0037</td>
<td>0.0073</td>
<td>0.0031</td>
</tr>
<tr>
<td>Overworld</td>
<td>R 0.93</td>
<td>0.67</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>p 0.00011</td>
<td>0.034</td>
<td>0.0062</td>
</tr>
<tr>
<td>Death</td>
<td>R 0.75</td>
<td>0.28</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>p 0.052</td>
<td>0.54</td>
<td>0.069</td>
</tr>
<tr>
<td>Flagpole</td>
<td>R 0.74</td>
<td>0.22</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>p 0.034</td>
<td>0.59</td>
<td>0.69</td>
</tr>
</tbody>
</table>

The summary of the results is as follows. There is a positive linear relationship between the pitch contours of M(ario) instrumentals and their linguistic interpretations in A(cappella) or S(peech) mode. Relative to the pitch contours of the speech renditions, the pitch contours of the a cappella renditions are closely related to the pitch contours of the
Mario instrumentals. With the exception of the Death and Flagpole tunes, the relationships between the pitch contours of all the music themes and their linguistic interpretations are statistically significant. The results also show that there is a positive relationship between the pitch contours of the interpretation in speech and a cappella modes, but the degree of similarity varies depending on the Mario instrumentals which is interpreted.

5.1 Linguistic interpretation of music involves acoustic mapping

The results of the acoustic analysis show that the pitch contours of the Super Mario theme music are similar to those of their linguistic interpretations, but depending on the theme music, the strength of this relation is either strong or weak. To explain the similarity, it would be useful to refer to the talking drum rendition of Yorùbá words. As reported in Akinbo (2019), Yorùbá drummers communicate by distinctively representing the pitch contours of tones in Yorùbá with their native drums. The results of the study show that there is a strong positive relation between the pitch contours of Yorùbá tones and those of their corresponding drum renditions. That the Yorùbá drummers musically encode the tones in Yorùbá is an example of sound imitation: “a process by which an individual either vocally or non-vocally generates sounds with qualities that reproduce elements of previously experienced sounds” (Mercado III et al. 2014: 39).

Comparatively, in the linguistic interpretation of Super Mario music and the talking drum rendition of Yorùbá words, there is a positive relationship between the pitch contours of non-vocal music and those of their corresponding linguistic forms. Based on this similarity, we can say that the linguistic interpretation of instrumental music, just as the drumming of Yorùbá words, is based on sound imitation or phonetic iconicity.

Despite the similarity between the linguistic interpretation of the Super Mario music and the talking drum rendition of Yorùbá words, there are certain differences. Tone-based speech surrogates do not encode segmental features (McPherson, 2019), but the vocal imitation of the Super Mario music contains segmental features. If we take into account the fact that vowel and consonant types affect the pitch value of a tone (Hombert, 1977; Whalen and Levitt, 1995; Whalen et al., 1999), the positive relationship between the pitch contours of the linguistic interpretations and those of the corresponding non-vocal music being strong or weak could be an effect of segmental features on the pitch value of a tone. When we compare the phonetic properties of the linguistic interpretation to those of the Super Mario Mario theme music, we observe that the music notes of the Super Mario theme music do not contain information about segmental properties (see Jackendoff and Lerdahl, 2006; Jackendoff, 2009). So, this begs the question: what determines the segment or morpheme that is matched to the musical patterns of the Super Mario theme music? In the next section, we focus on this issue.

Another interesting aspect of the results involves the correlation between the pitch trajectories of the interpretations in speech and a cappella modes. Although there is a positive correlation between the speech and a cappella of the musical interpretations, the strength of the relationship varies depending on the Mario instrumentals which is musi-
cally interpreted. In fact, the correlation between the pitch contour of the speech and a cappella for the Death tune instrumental is not statistically significant. This is in line with the hypothesis that the pitch trajectories of music melody in a tone language is not determined by language, but music can accommodate language when it is musically feasible (Schellenberg, 2012, 2013; McPherson and Ryan, 2018).

Before turning to the segmental properties of the vocal imitation, the discussion in this section is summarised as follows. Yorùbá gamers linguistically interpreted the Super Mario theme music by mapping the pitch contours of the music notes to Yorùbá phrases with similar pitch contours. Specifically, the gamers achieved this pitch-contour matching by modifying a suitable Yorùbá idiom or generating strings of words with similar pitch contours as the instrumentals.

6. Situational contexts contribute to linguistic interpretation of music

I account for the segmental or lexical properties of the linguistic interpretation of the Super Mario theme music in this section. For this purpose, we must refer to the issue of ambiguity in drum language. In phoneme-based speech surrogates, ambiguity is a current problem, and it is attributed to the drum language relying solely on tones of the source language (Ajayi, 1990; Poss, 2005; Pulleyblank, 2009). In speech surrogate research, there is a consensus that situational contexts resolve the recurrent problem of ambiguity in drum communication (Stern, 1957; Euba, 1990; Agawu, 1995, 2001).

The ambiguity of drummed speech and its resolution are clearly illustrated in the theme of “Oníbodè Lálúpọ̀n”, a poem written by the Yorùbá writer, Adébáyò Fáléti (Olátúnjí 1982: 2-5). The poem is about an event between an unattractive wealthy man called Oníbodè Lálúpọ̀n and a drummer who prefers to only play for wealthy people. When the drummer passes by the house of Oníbodè Lálúpọ̀n, he gifts money and palm-wine to the drummer. To acknowledge Oníbodè Lálúpọ̀n’s generosity, the drummer plays a tune on his drum (3a). Oníbodè Lálúpọ̀n interprets the drum tune as a praise tune (3b), but the detractors of the drummer tell Oníbodè Lálúpọ̀n that the tune is an insult to his unattractiveness (3c). When Oníbodè Lálúpọ̀n realises that the two interpretations of the drum tune are possible, he questions the drummer about the intended meaning of the drum tune. The drummer responds that his interpretation is of praise not insult. By taking into account the situational context of “gifting and appreciation”, Oníbodè Lálúpọ̀n accepts the interpretation of the drummer.
(3) Drummed speech in Onibodè Lálúpọ̀n

<table>
<thead>
<tr>
<th>a. DRUM TUNE</th>
<th>b. PRAISE INTERPRETATION</th>
<th>c. INSULT INTERPRETATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMMLLL</td>
<td>Mo jeun Ejìgbò</td>
<td>Ìwẹnu imàdò</td>
</tr>
<tr>
<td></td>
<td>'I ate Ejigbo’s food'</td>
<td>'Look at the mouth of the warthog'</td>
</tr>
<tr>
<td>MMMLLL</td>
<td>Mo jeun Ìwó</td>
<td>Ìwẹnu Ìsín</td>
</tr>
<tr>
<td></td>
<td>'I ate Iwo’s food'</td>
<td>'Look at the mouth of the minnows'</td>
</tr>
<tr>
<td>MMMMHMLHHHM</td>
<td>Mo jeun Onibodè Lálúpọ̀n</td>
<td>Ìwẹnu Onibodè Lálúpọ̀n</td>
</tr>
<tr>
<td></td>
<td>I ate Onibodè Lálúpọ̀n’s food</td>
<td>'Look at Onibodè Lálúpọ̀n’s mouth</td>
</tr>
</tbody>
</table>

Similar to how ambiguity is resolved in the poem, the gamers also utilised the situational contexts in the Super Mario game to determine the segments or words which are mapped to the theme music of the game. For example, the phrase “(s)he/it entered and cannot definitely come out” (4a) is mapped to the Flagpole fanfare theme which plays immediately Mario enters the castle. If we consider that the linguistic interpretation of the Flagpole theme involves mapping the sequence of tones H M H L L M H M to the music notes of the Flagpole theme, any phrase with the same tone melody could have been mapped to the music (4b-d). Mapping the phrase “(s)he/it entered and cannot definitely come out” to the Flagpole theme music suggests that the linguistic interpretation of the Flagpole theme is conditioned by the context of Mario “entering the castle and not coming out”.

(4) Possible interpretations of the Flagpole theme

Tone: H M H L L M H M

<table>
<thead>
<tr>
<th>Tone: H M H L L M H M</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Ò wọ́lè kòdè le jádé</td>
</tr>
<tr>
<td>b. Ò ta mì tètè lọ́ w’ágbo</td>
</tr>
<tr>
<td>c. Déwọ́lè àgbádo dára</td>
</tr>
<tr>
<td>d. etc.</td>
</tr>
</tbody>
</table>

Similar to the interpretation of the Flagpole theme, the situational context in the game is also utilised in the interpretation of the Death theme. For instance, the Death theme is played when Mario dies. Mapping the phrase “it is a dangerous game that you came to play” (5a) to the music notes of the Death theme correlates with mocking the death of Mario for playing a dangerous game. The theme of danger is also echoed in the linguistic interpretations of the Overworld theme (5b) and the Underworld theme (5c).
The interpretations of other Super Mario themes

a. **Death theme**
   Tone: M H H M H M
   erékéré lo wa sê ‘it is a dangerous game that you came to play’

b. **Overworld theme**
   Tone: H M H L L M H M
   tèbà làyé ko wá gbôbê l’órun ‘make èbà on earth, come to heaven for soup’

c. **Underworld theme**
   Tone: H H M M H H
   níbí lo ma kú sí ‘you are going to die here...’
   bóyá lo ma dé ’lê ‘(I) don’t think you will get home’

The Overworld theme is the main theme of the game, and it is the theme which is heard at the beginning of level one. The interpretation of the Overworld theme, which is “make èbà on earth and come to heaven for soup” (5b), is an adaptation of a Yorùbá idiom that is used as a death threat. Considering that the objective of the game is for Mario to save the Princess or die trying, the gamers possibly interpreted the Overworld theme as a warning or threat which is meant to create tension for the player-controlled character (i.e. Mario or Luigi) or more specifically the player. If Mario survives the first level of the game, the Underworld theme plays immediately at the beginning of the second level which is set in the Underworld. Interpreting the Underworld theme as “you are going to die here” is a threat to the player who escaped death at level one.

As mentioned in §2, the Underworld theme involves the repetition of the first six notes. Although the linguistic interpretation of the music melody and its repetition which is “I don’t think you will get home” are semantically different, they have the same communicative functions because both interpretations are death threats to the player. This suggest that two linguistic interpretations are possible for the same tone pattern inasmuch as each interpretation matches the situational context of music performance.

The summary of the discussion in this section is that the gamers utilised the situational contexts and visual imagery of the game in their interpretation of the Super Mario theme music. More specifically, the situational contexts of the musical performance assigns meaning to the vocal imitation of the instrumental music.

### 7. Formal account of musical meaning

The goal of this section is to present a formal account of how Yorùbá gamers interpreted the Super Mario theme music. For formal accounts of musical meaning, Patel (2008) identifies two approaches, namely semantic and pragmatic. In the semantic approach, the idea is that instrumental music can prime representations of meaningful concepts (Hacohen and Wagner, 1997; Koelsch et al., 2004; Steinbeis and Koelsch, 2011). Under the pragmatic approach, musical meaning is derived through contextual information and multimodal channels (Feld, 1984; Agawu, 2001; Patel, 2008; García et al., 2014; Neumeyer, 2015). In this
section, we will briefly discuss how these two approaches can account for the interpretation of Super Mario instrumentals by Yorùbá gamers.

First, we start with the semantic approach. In the semantic approach to musical meaning, musical features, such as pitch, tempo and rhythm, are capable of expressing affective and non-affective meanings. For instance, research on adult listeners has shown that faster tempo and high pitch prime affecting communication which are related to joy, happiness and pleasure and that slow tempo and low pitch prime the affective meaning of sadness (Scherer and Oshinsky, 1977; Kellaris and Kent, 1993; Gagnon and Peretz, 2003; Webster and Weir, 2005; Steinbeis and Koelsch, 2011; Eerola et al., 2013). Similar results are found in studies that focus on children music listeners (Dalla Bella et al., 2001; Mote, 2011).

The study of musical meaning has also been extended to non-affective meanings or concepts. To determine whether unfamiliar instrumental music could suggest semantic concepts, Koelsch et al. (2004) investigate semantic priming of words using unfamiliar instrumental music excerpts and sentences. Some musical stimuli in the study are iconic (resembling bird sound, sigh, etc), while other musical stimuli have stereotypical forms or style that are associated with particular words (e.g. devotion and caravan). The music-word pairing for semantic relatedness was based on ratings from a separate behavioural experiment. For semantic priming with sentences, a sentence such as “She sings a song” facilitates the processing of a semantically related word such as music, not sock” (Koelsch et al. 2004: 302). In the experiments, the participants listened to a musical excerpt or a sentence, then they were visually presented a semantically related or unrelated word. After the visual presentation, the participants were asked whether the prime and the target were meaningfully related or not by pressing one of two buttons. Koelsch et al. (2004) utilised a neural signature of linguistic semantic processing known as the N400. The results of the study show that the participants categorised 80% of the target words correctly. The N400 priming effect in the music domain did not differ from that of the language domain. Following from this, Koelsch et al. (2004) argue that music and language have ability to activate semantic concepts. In general, the semantic approach predicts that the pitch values of a music note should have a specific meaning.

There is some evidence to suggest that music can prime semantic concepts, but as Patel (2008: 334) notes, it does not mean “that music has a semantic system on par with language”. For instance, unlike human languages where vocal or visual signals without any external stimuli can have a semantic meaning, an instrumental music without a matching word or any other stimuli has not been shown to elicit a semantic processing (see Hacohen and Wagner, 1997; Steinbeis and Koelsch, 2011; Koelsch, 2005). Of course, the prediction of the semantic approach might not hold in a music tradition with musical speech surrogate like Yorùbá. For instance, a specific pitch value or tone might be associated with words that refer to opposing concepts such (e.g. happiness and sadness). If music really elicit a semantic processing, the music notes in the Underworld theme of Mario and its repetition would have been assigned the same meaning. Considering that Yorùbá gamers utilised their cultural background and visual information of the game in their interpretation of the background music, it is highly possible that the meaning of the instrumental music
in Koelsch et al. (2004) is derived from coupling music with the visual signals, not music alone. Following from this, the semantic account of musical meaning is not compatible with the interpretation of Mario instrumentals by Yorùbá gamers.

As an alternative to the semantic approach, Patel (2008) argues for a pragmatic approach as a more efficient way to studying musical meaning. Central to the pragmatic approach in Patel (2008) is the coherent structure of discourse. For instance, the sentences or clauses in (6a) are coherent because they are related in systematic ways and form a unified meaningful whole. In line with the maxim of relevance (also relation) in Grice’s Cooperative principle (Grice, 1975; Leech, 1983), the second utterance is relevant because it contributes to the communicative goal.

(6) Linguistic discourse (from Hobbs, 1979)
   a. Coherent  [John can open Bill’s safe]. [He knows the combination].
   b. Incoherent  [John took a train from Paris to Istanbul]. [He likes spinach].

By assuming that coherence relations between clauses in linguistic discourse are analogous to coherence relations between phrases and themes in musical discourse, Patel (2008) suggests that investigating coherence in musical discourse could offer more valuable insights on musical meaning. The pragmatic approach predicts that the internal structure of a musical piece, such as the number of musical notes, should have an effect on musical meaning. The results of the experiments in Poss (2012) can be considered an argument in support of the pragmatic approach in Patel (2008). For example, Poss (2012) conducted an experiment on the cognition of raj nplaim music, which is a musical speech surrogate of Hmong people. The results of the experiment shows that the participants were able to match Hmong words to the musical melodies of the raj nplaim instrument, but they found it easier to match phrases to longer musical melodies. That the participant performed better on longer musical phrases could be an effect of coherence. The pragmatic approach of musical meaning in patel also has a similar limitation as the semantic approach because it suggests that musical meaning is solely dependent on the internal structure of music. Considering that the Yorùbá gamers used the context of musical performance in addition to the pitch contours the instrumental music, the prediction of the pragmatic account is also not compatible with the interpretation of Mario instrumentals by Yorùbá.

In other approaches which can also be classified as a pragmatic approach, musical meaning is derived from both the internal structure of musical discourse and the situational contexts relating to musical performance or consumption (Feld, 1984; Neumeyer, 2015). In his work on music communication, Feld (1984) argues that musical structures exist in social constructions and they have meaning through social interpretations. To account for musical meaning, Feld (1984) proposes a music communication model which is schematically represented in Figure 16. In the model, music communication process involves two dynamically linked components, namely sound objects and interpretive moves. The interpretive moves contain five elements: (i) a locational element which relates a sound object to an appropriate range within a subjective field of like or unlike items/events; (ii) a categorical move which relates sound to things (e.g anthems and patriotic songs); (iii) an
associational element which relates the sound event to a visual, musical or verbal imagery; (iv) a reflective element which relates a sound object to personal and social conditions, and experiences where things like and unlike the object can be heard; (v) an evaluative move which involves an affective meaning (Feld 1984: 8). While all or some of the five elements might be utilised in the interpretative moves, the element utilised in an interpretative move must interact with the identity of the listener, an expressive ideology and world sense coherence.

The interpretive moves proposed in Feld (1984) can account for the interpretation of the Super Mario themes. For this purpose, we refer to the discussion in the previous section. The discussion suggests that the Yorùbá gamers used their cultural experience, pitch imitation and situational contexts of the game in their interpretation of the Super Mario instrumentals. Drawing on their native cultural experience in the interpretation of the Super Mario theme music can be considered a reflective interpretive move. That the gamers imitated the pitch contours of the instrumental music can be accounted for with a locational interpretive move, which involves relating a sound object to an appropriate range within a subjective field of like or unlike items. In this case, the theme music of the game is possibly likened to the talking-drum music in Yorùbá. Given that the associational
interpretive move involves relating sound event to a visual, musical or verbal imagery, the use of the visual events in the game for the interpretation of music can be considered an effect of the associational element.

In sum, the music communicative model in Feld (1984), which can be considered a pragmatic approach to musical meaning, accounts for the interpretation of the Super Mario themes by Yorùbá gamers.

8. Summary, Discussion and conclusion

This study has investigated the linguistic interpretation of Super Mario musical themes by Yorùbá gamers. The results of the study suggest that the instrumentals were interpreted through multimodal channels or moves which might be unordered. The first step involves one-to-one mapping between syllables and music notes of the instrumentals. To determine the tone of the syllables, the second move involves mapping the pitch contours of the music melodies to Yorùbá phrases with similar pitch contours. However, the degree of similarity between the pitch contours of the music melodies and their linguistic interpretations varies depending on the music theme. In the third move, the choice of segments or morphemes which is mapped to the music is determined by the situational contexts of the musical performance. In this case, situational contexts include the visual events surrounding the instrumental music in the game, non-virtual events related to the visual events in the game and the cultural background of the listeners (i.e. the Yorùbá gamers).

That Yoruba gamers mapped the pitch contours of the music melodies to the vocal forms is in line with previous studies on vocal imitation of instrumental music (Lemaitre et al., 2016; Mantell and Pfordresher, 2013; Mercado III et al., 2014; Wisniewski et al., 2013). As reported in studies on Yorùbá talking drums, the communicative capability of the speech surrogates is based on mapping the pitch contours in speech forms to musical melodies (Villepastour, 2010, 2014; Akinbo, 2019). This suggests that the musical speech surrogates and the linguistic interpretation of music involve pitch imitation in Yorùbá.

The most important contribution of this work is the role of contextual information in the interpretation of unfamiliar instrumental music. That the Yorùbá gamers used contextual information in the interpretation of the instrumental music is an argument in support of the pragmatic approach to musical meaning (Feld, 1984; Patel, 2008; García et al., 2014; Neumeyer, 2015). While most of the studies on musical meaning or the linguistic interpretation of music are based on experimental studies, the linguistic interpretation of the Super Mario music by Yorùbá gamers presents natural evidence in support of the linguistic interpretation of instrumental music.

Given that the linguistic interpretations reported here developed naturally and conventionalised with many iterations of generational overturns, there are certain limitations of the present study. For instance, we are uncertain whether, in addition to the factors mentioned in this work, other factors play a role in the linguistic interpretation of the Super Mario music. In a controlled environment, an important issue to address in future studies is how the context of music performance contributes to the linguistic interpretation of
instrumental music. For such studies, embedding the musical stimuli as the background music of a specially designed video game could be promising. Such studies will not only contribute to the fields of linguistics, psychology and music, they will contribute to the emerging field of video game music, where there is a consensus that background music of a video game increases the level of “cognitive immersion” or “being in the game” (e.g. Zehnder and Lipscomb, 2006; Grimshaw, 2008; Nacke et al., 2010; Sanders and Cairns, 2010; Fu, 2015). The use of cultural experience and situational contexts in the interpretive moves by the Yorùbá gamers is possibly because music in African tradition is functional, linked to dance and externally motivated by social and musical contexts (Agawu, 2006, 2001). Future research would investigate whether functional and non-functional music are interpreted differently. It would be interesting to investigate whether there is a meaningful distinction between interpreting music on the fly in the moment and a conventionalized pattern as is the case here.

It bears mentioning that only some portions of each music theme is linguistically interpreted. For future research, it would be interesting to investigate whether this has to do with the rhythmic properties of the portions which are not linguistically interpreted. If this is the case, we could investigate whether this has to do with the rhythm of these specific portions of the music themes in relation to the rhythm of speech (see Patel and Daniele, 2003; Patel, 2006, for comparative studies on rhythm in speech and music).

To conclude, Yorùbá gamers linguistically interpreted instrumental music by vocally imitating the pitch contours of musical melodies and mapping a lexical syllable to each musical note. The assignment of morpheme to the vocal imitation is through the use of the situation contexts surrounding the musical performance and the cultural background of the listener. Considering that many studies on African oral poetry do not employ video game, I recommend the use of video-game data for studies on music-language connection.

References


