PRESENTACIÓN
EQUIPO EDITORIAL
INFORMACIÓN PARA LOS AUTORES
CÓMO CITAR TRANS
INDEXACIÓN
CONTACTO

Última publicación   Números publicados

< Volver   TRANS 13 (2009)

Synthesizing Race: Towards an Analysis of the Performativity of Vocal Timbre

Nina Eidsheim

Abstract

Vocaloid is a vocal synthesis software package that "sings back" any pitch and word combination entered by a user, impersonating a singer with a designated sex, age and race. Lola and Leon, the first pair of "singers" designed, were introduced as "generic soul-singing voices." Investigating vocal timbre as a cultural artifact, I look at the processes by which audience connect specific vocal sounds with particular ideas such as race and gender. Such reification of notions of race through vocal timbre is circular: audiences join sounds with concepts; (live or digital) performers respond to these sound/concept compounds, and in turn confirm the listeners' linkages. Thus an analysis of timbre as an inner choreography is necessary to begin to map and denaturalize the connection between vocal timbre and race.

Key words: Music and technology, performance studies, performativity, semiotics, Soul, Vocaloid, vocal synthesis, vocal timbre, voice

Resumen

Vocaloid es un programa de síntesis vocal que "canta" cualquier combinación de alturas y palabras dadas por un usuario. Al hacerlo, el programa imita la edad, raza y género sexual del cantante según sean designadas. Lola y León, la primera pareja de "cantantes" designados, fue presentada como "voces genéricas de cantantes de Soul". Al estudiar el timbre vocal como un artefacto cultural, este trabajo observa los procesos por medio de los cuales la audiencia conecta cierto tipo de sonidos vocales con ciertas ideas sobre raza y género. Esta reificación de la noción de raza por medio de timbre vocal es circular; la audiencia asocia sonidos a conceptos; los intérpretes (tanto en vivo como digitales) responden a este tipo de relaciones binarias sonido-concepto y reafirman las asociaciones de la audiencia. Así, es necesario un análisis del timbre como una coreografía interior para empezar a mapear y desencializar las conexiones entre timbre vocal y raza.

Palabras clave: música y tecnología, estudios de performance, "performatividad", semiótica, Soul, Vocaloid, síntesis vocal, voz.

Mr. Darden: "The second voice that you heard sounded like the voice of a Black man; is that correct?"

California v. Orenthal James Simpson (1995)[1]

Introduction

Investigating the complex of perceptions surrounding vocal timbre can be revealing. When social categories associated with vocal timbre, such as gender and race, are identified as inherent characteristics of individuals,[2] they function as subtle gatekeeping practices which control access to social positions and their attendant societal benefits. Institutions (and sites of power in general) replicate themselves and seek to control change. Practices such as the reification of vocal timbre may offer insight into that process. Studying these practices can be challenging because: 1) as previously stated, the mechanisms of reification function in extremely subtle ways within existing (and often oblique) structures of power; 2) the embodied nature of the voice causes timbre to seem immanent to individual bodies, rather than encultured; and 3) it has historically been imagined that the voice discloses intimate and truthful information about identity.

The paucity of available research on these issues may be attributed to a more complex problem than mere analytical difficulty. Even *interest* in understanding such practices may be limited by the ways in which inequities are often hidden by media attention and its focus on a small number of successful

resonant in his vocal timbre,[3] it is clear that the voice is generally regarded – not only colloquially, but also legally – as evidence for inherent racial differences. Acceptance of the voice as a basis for legal decisions reveals that, in general terms, it is considered to be the unmediated sound of a poorly-defined "essence" of the body – and that this "essence" is defined through, for example, racial or gendered terms.

If the relationship between vocal timbre and race is not an essential one, but if there seems nonetheless to be some relationship, although unclear, between the body, vocal timbre and identity, how can we explain it? Exploring this conundrum through analysis of the vocal synthesis software Vocaloid will bring two issues sharply into focus: first, a critique of the faulty assumptions upon which vocal timbre, as described, is generally read; and second, a proposal that the relationship between timbre and the construction of identity may be understood more accurately through notions of performativity.

In what follows, I draw on the work of John Baugh (2003), Nancy Niedzielski (1999), D.L. Rubin (1992), and George Lakoff (1987) in order to develop an understanding of the processes by which the voice is interpreted and categorized. Stuart Hall's (1980, 1986) work on articulation clarifies the fluid relationship between sound and meaning, and brings to the fore the active role played by both performer and listener in the flux of this relationship. While my work rests on John Shepherd and Peter Wicke's (1997), Richard Middleton's (1990, 2003, 2006), and Phillip Tagg's (2008) previous work on music semiology and analysis, this study also proposes an analytical framework that takes the physical body of the singer into account. My effort is informed by literature from performance studies. I develop a theory of articulation through performance, and outline the *performativity of vocal timbre*, in order to arrive at an understanding of the context-dependent processes of meaning which have previously been interpreted as the essential sounds of the body. This study reveals two crucial points: that a person's vocal timbre cannot be entirely unmediated; and the (many possible) meanings we derive from any given vocal timbre are not immanent. It is the space between the not-entirely-unmediated and the not-immanent––the *performed articulation*––and its impact on the physical body of the singer (whether this "body" is constructed through a computer application or belongs to a "live" singer) that is mapped by this paper.

Vocaloid

The commercially available vocal synthesis software Vocaloid was first introduced to the American market in 2004 at the National Association of Music Merchants (NAMM)[4] in Anaheim, California––one of the most important annual music industry tradeshows. The application received enormous attention, garnering several music and technology journal awards. A *New York Times* review published after the 2003 European release[5] (in anticipation of the NAMM show) hailed this Yamaha synthesis method as a "quantum leap" in human voice modeling. The attention and praise Vocaloid received from industry professionals and journalists was, for a time, consistently high.

Vocaloid is described by its developers, the British music software company Zero-G, as a "vocal-synthesizing software that enables song writers to generate authentic-sounding singing […] by simply typing in the lyrics and music notes of their compositions."[6] The software is further compared to a "library" in which the user may imagine the different voices as "vocal fonts." (See Figure 1 of screen shot.) The user enters pitches and durations on conventional staff paper in one application setting, or by playing the piano interface (or a connected midi piano) in another setting. Lyrics are added as the user types them in, in correspondence with the notes; melody and words are then sung back by the voice the user selects. This process is roughly comparable to typing words into a text document and having them read back by text reader software; however, unlike a conventional text file reader, Vocaloid assigns pitch and duration to each word based on user input. Also, each pitch-duration-word compound may be treated with added vibrato, envelope, attack, dynamics, and so on. In the same way that a few mouse clicks will change the font type in a word document, a Vocaloid song may potentially be sung back by any of the available voices.[7]

http://www.zero-g.co.uk/media/images/vocaloid_screenshot.jpg   *Figure 1: Screen shot of Vocaloid*

Each Vocaloid voice is made up of thousands of samples recorded by a single singer.[8] Together the samples represent about 3,800 possible vowel and consonant combinations found in the English language. Each original singer recorded sixty pages of scripted articulations (e.g. [pel, pep, lep], etc.) on three different pitches, which were then manually trimmed into precise samples. The fact that this process required eight hours of recording per day for five days may offer an idea of the sheer volume of these combinations.

One of the main challenges in creating software that sings words is the translation between the *spelled word* entered by the user and the actual *sounded phoneme*. For example, the word "Philadelphia" begins with a phoneme similar to the word spelled "fish." One of Vocaloid's tasks is therefore to choose the recorded phoneme that corresponds to the written word; in the above example both words, although spelled differently, begin with the same phoneme. The synthesis procedures used in Vocaloid were developed through a collaboration between Pompeu Fabra University (Barcelona) and Yamaha. Vocaloid's synthesis,[9] using the system described above, combines the recorded phoneme samples into a seamless string forming words sounded in melodic sequences. In electroacoustic music terms, Vocaloid may be considered *hybrid vocal synthesis* in that it uses basic sonic material from the phoneme recordings (sound samples are not used in "complete" sound synthesis). The application relies on traditional synthesis techniques in order to combine and alter the sounds of the samples.[10]

Prior to Vocaloid, vocal synthesis applications were described in terms of their technological advancements and their advantages as powerful sound synthesis tools. In contrast, Zero-G has offered each vocal font not merely as a synthesis application, but as a *singer*. Different versions have been marketed in varying degrees of detail, but all have been given Christian names. For example, the first two voices to be released were given the names Lola and Leon. A third edition was called Miriam, which was the name of the singer, Miriam Stockley, who provided the voice samples for the synthesis.[11]

Besides names, the Vocaloid applications were each assigned a personal profile, from the specific (identification with a particular singer, such as Miriam Stockley) to more general categorizations in terms of, e.g., genre and gender. Lola and Leon were marketed in the latter manner: Zero-G describes them as "the world's first virtual" male and female "soul vocalists," whereby the gender, genre, and possible applications of each voice are highlighted. In contrast, Miriam is described in personal and concrete terms as "based on British singer[12] Miriam Stockley's voice." Rather than closely identifying this application with a genre, an emphasis is placed on the notion that "[her] voice is pure and suitable for the current synthesis engine,"[13] and that she is a "virtual vocalist."

The profiles as a singer rather than software, was also carried through in the packaging. While the (highly problematic) images on the box of Lola, Leon, and Miriam (Figures 2-4)[14] to a varying degree depict a person, in contrast an older synthesis software package such as Cantor (Figure 5) highlights the technical aspect with an anatomical drawing of the vocal tract and sound waves.

Where Lola and Leon were seemingly created with an image of anonymous backup vocalists in mind and where even the same images was in fact recycled by being re-tinted and reversed, Miriam originated in the idea that the user may be offered access to the familiar face and voice of a popular singer. At the time of the Miriam application's creation, Miriam Stockley was known to a broad audience through UK chart placement with the song "Only You" (1991) and the 1995 Karl Jenkins project "Adiemus", recorded for a Delta Airlines commercial. The music was so popular that it was released on the album *Song of Sanctuary* with the London Philharmonic Orchestra. Subsequently it has been used in movie tracks (e.g. *Invisible Children*) and in several other commercials.[15] Rather than representing a *genre* (as in the case of Lola and Leon), it is the unique artist Miriam Stockley—the voice of an individual artist—who gives identification to the Miriam Vocaloid singer.[16]

http://www.zero-g.co.uk/index.cfm?articleid=800   *Figure 2: Lola*

http://www.zero-g.co.uk/index.cfm?articleid=799   *Figure 3: Leon*

http://www.zero-g.co.uk/index.cfm?articleid=805   *Figure 4: Miriam*

http://www.guitarcenter.com/VirSyn-CANTOR-2-1-The-Vocal-Machine-Software-701957-i1176772.gc

Figure 5: Cantor

Users' Perceptions of Lola

Despite Zero-G's comprehensive efforts to present a black soul singer, many of Lola's users did not hear her voice as a soul voice, and/or as black. User RobotArchie wrote on Zero-G's internet message board: "Do we have a British soul singer with a Japanese accent who lisps like a Spaniard? Eesa makea me tho unhappy...." Heatviper chimes in: "Hello... I think Lola works great for mondo/mournful /giallo morricone style tracks using vowels....wordless soulful vowels are nice." Jogomus asks for advice: "My Lola sounds a little bit like a "big Ma" - what can I do, [so] that she sounds a little bit neutral?" HK suggests lowering the "Gender Factor" value (Vocaloid:User).

In addition to comments such as the above, users reported that an unexpected and problematic accent emerged during their implementations of Lola's voice, an accent which became difficult for Vocaloid to explain. The head programmer reported subsequent online exchanges with users wherein he, with convoluted technical explanations of the synthesis method, tried to obfuscate the fact that some users found the performer's pronunciation strange or unexpected in relation to the anticipated black soul singer's voice.

After the first glowing reviews and awards, it could be gleaned from reading the Vocaloid user forum that general reception was less uncritically accepting. User discussion centered on the appropriate genre in which to use the software, precisely the aspect of the product that Zero-G had worked most diligently at defining. Whereas the *New York Times* reviewer was interested in Vocaloid's potential to revive the voices of famous singers such as Elvis (Werde, 2003) through extracting existing sound samples from recordings and patching them together with the new synthesis method, users of the software took a more practical approach. They listened to the applications and thought about what these voices sounded like – and they discovered that the sound failed to match their conceptions of the black soul singer's sound and the product advertised by Zero-G.

Because the sound of a standard vocalist within the soul genre is well defined, it is safe to assume that both Zero-G and Vocaloid's users possessed reasonably similar ideas about the ways a soul singer's voice should sound. However, an apparent gap emerged between the product Zero-G wished to sell and what its users, or at least those who participated in the user forum, experienced. This gap was articulated by the differences between the software profile created by Zero-G—the composite of sound, visual representation, textual description and genre reference—and the experiences about

Signification Through Vocal Timbre

Some basic semiology may prove useful in understanding the depth of the disparity between what Vocaloid wished to produce and users' reactions to what it did produce. The *signifier* is that which has physical existence, the sign materialized, *as we perceive it*. It is the figure in the photograph or the sound in the air. The *signified* constitutes our mental concept of the signifier: the meaning that physical existence takes on when mediated by culture. For example, all members of the same culture share the perception that a particular shape featured in a photograph is a car, and that a particular sound in the air is a school bell. On this connotative level the *sign* is defined as the associative, nonlinear sum of the signifier and the signified.

While both denotative and connotative levels of signification describe the meanings conveyed by signs, drawing a distinction between the two *types* of meaning that they can convey is important for a thorough analysis. We can read both denotative and connotative meanings from the same sign. So, using a sonic example, the denotative meaning of a sound might be an idea of the sound itself––e.g. the recognition that this is a human voice singing B flat above middle C. The way in which that sound is recorded (the quality of the voice)––e.g. with a lot of reverb, low-fi, or very brightly––can constitute its connotation.

Additionally, signs are read within three orders of signification, or three levels of meaning. In the first order, the sign is read as self-contained, i.e. the sound is heard as a B flat above middle c sung by a voice (as opposed to, for example, the same pitch as played on a piano). In the second order of signification the possible meanings of that sign are read not only within the available range of concrete meanings, but also within a range of cultural meanings – which are not derived from the sign itself, but arise from the ways in which a given society uses and values it (signifier and signified). So, if the B flat is played back with filters that cause it to sound like a low-fi recording, its connotative meaning may be an aesthetic that posits itself as DIY and counter-commercial; and the same B flat's connotative meaning will likely change with the quality and amount of reverb applied to the recording. With a large amount of reverb, the B flat's connotative meaning could be a cathedral, or at least a religious space. It is in the third order of signification that all of the culturally-dependent ranges of meaning within the second order cohere to create a more general picture of that world. In the third order our B flat may be understood as a building block of the hegemonic Western diatonic system, and our digital filters as elements of a commoditized music system wherein the consumer may purchase whatever "acoustic space" she wishes. In the third order of signification the myth (to use Barthes' term) that operates as the organizing principle for the values or ideology for a given community reflects the broad principles to which a community adheres, and through which it coheres. Such mythology derives from both individuals and social institutions.

Reasoning Using Categories

These myths inform a variety of cognitive impulses. George Lakoff (1987) proposes that when we reason, we reason not only in terms of individual things, but also in terms of categories. Our entire world is divided into categories, and we ascribe to them a real existence and an inner coherence; to understand this process of categorization is thus crucial to comprehending the ways we reason. Lakoff suggests that categories are derived from both human experience and imagination, and that sometimes items belonging to the same category do not, in effect, possess common qualities (8). Operating within such a framework, then, we may trace the processes whereby a category that assumes an inner coherence between a particular vocal timbre and a black body is reified, and we may come to understand the assumptions which, in the case of Vocaloid, gave rise to a perceived direct connection between a black body and a soul vocal style and timbre.

Lola and Leon were explicitly labeled black soul singers by Zero-G. It is, of course, correct that the historical and social circumstances from which the genre arose were deeply involved with African American communities and culture. The term arose with black American gospel groups, e.g. the Soul Stirrers, in the 1940s and '50s, and subsequently became an umbrella term used to label black American popular music. Even Billboard adopted the term "soul" for its previously named "rhythm and blues" chart. As new music genres––such as funk and disco––which were quite unlike the music that had first been identified with soul gained chart success, Billboard found the term inappropriate for the category that, generally speaking, contained black American popular music of the day. Accordingly, the chart category was renamed "black music."

The soul genre today features an instantly recognizable vocal delivery modeled after the style of singers such as Clyde McPhatter, Ray Charles, James Brown, Otis Redding, Aretha Franklin, and Stevie Wonder––all black artists and singers whose music was shaped by their experience and life in African American communities (often with a close relationship to gospel music).

However, while there are many and close connections between the sounds of soul singing and black community and culture, the assumptions that (a) the sound itself is essentially black and (b) these connections indicate that African Americans express an essential blackness through their vocal timbre result primarily from the way in which race, as an organizing principle of American society, has given rise to a category which bundles together a particular body and a particular vocal timbre without real conceptual coherence– and a semiology which relies on these premises will necessarily be flawed. It was upon precisely such an incoherent category that the semiology which gave rise to Lola was based.

With the aim of creating two soul voices, Zero-G undertook a process akin to "reverse engineering." Following a chain of reasoning at the end of which soul equaled blackness, they chose two singers for

as "black and English", and offered that "he is a lovely guy as well...". About Lola, I was told that she is also "black."

It is through users' responses to the voices that the cracks in the producers' assumptions regarding an innate connection between blackness and genre appear. The programmer I talked to was surprised at the reactions of software users such as the contributers to the online forum quoted above. In an aside during our discussion about these reactions the programmer mentioned that the Lola singer was from a Caribbean background, but that she was often in demand as a studio singer for soul material since she sounded idiomatically like a soul singer.

Because Zero-G assumed that a soul sound would be emitted from any black body, they chose a black body to provide the sound samples. But when the Lola singer sang pure syllables outside the soul music context, her origin in the Caribbean––and thus an accent atypical for soul music––was recorded. In assuming an essential relationship between a black body and the soul sound, Zero-G assembled Lola using pieces that failed to add up to what we know as soul.

Users' rejection of Lola as a soul voice shows us that a vocal sound that we recognize as soul is not the essential sound of blackness which any black vocalist will automatically inhabit; instead it is comprised of a particular vocal delivery and timbre (with an indisputable origin in African American culture). In semiological terms, the sound with which the users were presented signified, on the first level, the sound of a voice; on the second level, a particular accent which users may not have precisely identified, but which they heard as different from the timbre they associated with soul. Therefore, because the users had been presented with the profile of soul voice, on the third level of signification they defined the voice in opposition to that image.

The Relationship Between Vocal Timbre, Body, and Race

If the example of Vocaloid reminds us that the relationship between a given body, its race, and its vocal timbre (and any vocal timbre we may recognize as representative of a genre) is not an essential one, what then is the relationship between the body and vocal timbre? The vocal timbre that arises from a body is a sound that is, whether or not the singer is aware of this process, carefully constructed. Such processes of construction may take place without the singer's awareness, or the process may be very clear. The particular vocal timbre adopted by each person through daily speech and singing activities exemplifies a situation in which the processes of construction can take place unnoticed. Many of the iconic soul singers mentioned earlier adopted the vocal timbres, for which they are loved and emulated, without deliberate effort; instead their vocal qualities were gradually shaped through the music-making they engaged in on a daily basis. A singer may also become skilled at producing a particular vocal timbre through conscious efforts such as voice lessons (almost always the case in classical vocal styles) or listening to recordings and imitating their vocal mannerisms. When a singer deliberately learns a vocal style and timbre, the creation of that timbre becomes transparent. For example, in the case of a classical vocal sound, most singers will be aware that the throat is, in general terms, more open than in most other vocal styles. In contrast, when the vocal style is adopted gradually through everyday vocalizing and music making, the singer and the audience may be less aware of the physical parameters of the style.

Whether that process is a seamless part of the singer's everyday musical life, or whether it is learned through formal voice lessons, the singer's body performs just a single subset of *a range of timbres it is possible to produce given her vocal apparatus*. The fact that most people, without questioning the logic of their conclusions, will readily draw correlations between a vocal timbre and the singer's so-called race is a symptom of the "standardization" of the concept of race in a given society – and, as I remarked in my introduction, the reification of the body and voice which takes place during this process. When a person is identified by the sound of her voice as African American, the sound of that voice represents the vocal community to which the singer belongs, or in which she desires to mark herself as a participant, rather than the essential sound of her body. That is, the correlation of such vocal communities with race, ethnicity or class is not inherent; instead it is a symptom of the divisions that are important in the society – and it is the *performance* of these divisions.

The concept of *articulation*, Stuart Hall (1986) has proposed, describes a point of connection between *two independent parts*, a connection that can be broken and established, carrying its own distinctive implications – as, for example, performing a presumed connection between sound and race. This notion captures Zero-G's connection between the concept of the soul sound and a black body. There is no direct correlation between the two; the articulative connection is forged in a listener's mind between two independent parts such as a sound and a racialized body. When the black body is assumed to be synonymous with a soul vocal timbre a *performed articulation* (rather than an inherent meaning of that vocal timbre) takes place.

Rather than being manifestations of essential physical states, then, these timbral indices may be more accurately considered as a set of *inner choreographies*––movements that create internal physical configurations that give rise to a timbral identity. Timbre, following this argument, is a physical configuration, and the resultant sound merely a confirmation that this internal shape has been performed. The relationship between vocal timbre, the body and race is a performed articulation connecting independent parts, rather than an expression of an essential relationship. But because the choreography that engenders timbre is internal, timbre has historically been considered the inherent sound of a body. Vocaloid's construction, and users' rejection, of Lola as a soul singer, however, offered a unique opportunity to examine the complex of perceptions surrounding vocal timbre and race. And it demonstrated that it is the performed articulations that bind a sound and its meaning together.

Listening is not an isolated, private matter. Therefore the audience's performed articulation of the meaning and value of a particular vocal timbre may affect the consequent work of the singer. One of my longterm projects is an ethnography of vocal students and teachers within the classical genre (Eidsheim 2008:30-66). The study concludes that the listening practices of many teachers are unconsciously framed by colonial and post-colonial attitudes toward race and ethnicity. This often-unconscious belief in racialized bodies steers listening – and with this aural compass the listener detects racialized vocal timbres, or finds a lack thereof unsettling and out of character. For example, a teacher who consciously or unconsciously believes in racialized bodies will often perceive a "Korean" vocal timbre in the voice of a student who appears to be ethnically Korean (see Eidsheim 2008: 28-29; 47-49). As we have established, soul vocal timbre – although clearly connected to the cultural history of African Americans – is not an essential sound of African American bodies. Similarly, a vocal timbre associated with classical vocal training in Korea is intimately tied to the geographical area and its people, but is not the essential sound of a Korean body – although a Korean singer, or another singer with a different ethnic background trained within the Korean tradition, will probably adopt its signature vocal timbre, while a singer of Korean origin (or any other singer of any other origin) who has not been trained within Korean tradition will likely not adopt its timbre.

In this case slippage exists between a particular vocal timbre and a tradition of singing which may be likened to a national school of singing (for example, Italian, Germanic, French, Slavic, Nordic), in the classical vocal world's sense. Various factors (unconnected to the range of vocal possibilities inherent in the bodies of Korean nationals or ethnic Koreans), including the influence of the native language's diction and regional or national aesthetic preferences, have given rise to what people recognize as the Korean classical vocal sound. I would like to draw attention to the fact that what has come to define the "Korean" sound is not race or ethnicity, but the standard vocal training available within the geographical area of South Korea. The slippage which takes place in a situation where, say, an American teacher expects to hear a Korean sound from an ethnic Korean who has not grown up in Korea (or is not knowledgeable about the Korean classical vocal ideal or the Korean language) illuminates listeners' articulations between a defined vocal timbre and a racialized body. This type of assumption (i.e., the assumption that any black body would produce a soul sound) underlies Zero-G's construction of Lola's sound.

Signification through vocal timbre presents a unique situation, in that the singer and his or her body are very flexible and sensitive to how they are received and perceived. A singer can easily, and often does, follow and change according to these perceptions. This has made it very difficult to conceptually separate the racialized body and vocal timbre. When a voice teacher hears a student through a perceptual framework of Koreanness, and the sound does not exhibit the essential attributes of a Korean body, but it also does not necessarily possess the timbre which would result from Korean vocal school training, and the teacher clearly hears a Korean vocal timbre, what is at play?

What takes place in this scenario is an articulation on the part of the teacher, which *connects* the meaning or value assigned by the teacher to, for example, Koreanness or blackness with the student's vocal timbre. When a student receives positive feedback regarding a particular articulation, she may often respond by producing the timbre expected from her. Thus a singer's vocal apparatus slowly adapts to the new sound required from it, and begins to take that form – and because of this continuous conditioning the vocal body (constituted by all aspects of a singer's physicality that are involved in and shaped by vocal engagement) begins to produce the timbre the teacher/listener believes she heard from the beginning and is merely fine-tuning. As a consequence of this process the voice student may, in the end, create what the teacher considers to be a Korean timbre, thus confirming the teacher's timbral expectations. In other words, any singer could, with proper practice, learn any given timbral category.

In summary, articulation through performance is the point at which the discursive impresses upon the corporeal and can, in effect, alter the corporeal. And because articulation through performance is a joining, corroboration or intensification of the discursive and material form, it may also engender a disconnection between the discursive realm and vocal timbre in the external world. Thus articulation through performance may both join and disengage a concept and a material form. This ability inheres in the performance of both singer and audience.

Vocal Timbre: a Sonic Phenomenon?

This work raises a more general question: why has timbre in general, and vocal timbre in specific, been so resistant to in-depth analysis within Western thought? One reason may be that timbre has historically been considered an exclusively *sonic phenomenon*—a component of sound only. Within such a framework we listen to the timbre presented to us and analyze it through devices such as spectral analysis, which measures sound waves (different levels of pressure transmitted through air). We investigate the nature of sound as if it is emitted in its final form.

If we consider vocal timbre as merely the sonic material presented to us, it seems that we unknowingly take shortcuts and assume that we hear the unmediated sounds of bodies. That is, if we consider timbre as a sonic phenomenon only, the shaping of the vocal apparatus, rather than being understood as a conscious or unconscious performance, is heard as the inherent sound of a body.

By considering timbre instead as the sound that results from the *vocal body—the vocal apparatus as it is fashioned through repetition of particular sounds, rather than the inner structure of an essential phenotype*—we may come to the realization that timbre is actively *shaped,* rather than passively projected. In essence, each part of the body that participates in the creation of vocal sounds (vocal tract, torso, tongue, mouth cavities and so on) has been actively fashioned. I term both this active sculpting of the vocal apparatus, and the shaping that takes place on the fly, *the performativity of*

a tale of initiation into the form's signature *sound,* wherein the (female) singer must go to the mountains, find a huge waterfall and sing against it with the aim of outdoing it in volume. The singer must go on singing until she spits blood. By severely bruising her vocal chords she has forever altered their physical shape, and thereby their sound.

 This p'ansori body is her new *vocal body*. The practice of p'ansori exemplifies a highly conscious *sculpting* of the vocal body with the aim of producing a particular sound.

Vocal timbre has also proved theoretically elusive because the feedback loop between audience expectations and singers' accommodations is invisible, and timbre seems to confirm belief in essential sound. When the voice – through, for example, visual cues – is perceived in a particular way, the singer may react by favoring sounds that garner positive feedback, for example, in the form of recognition and work opportunities. By reacting to teachers' and audiences' expectations and preferences, the singer's *vocal body* is shaped to emit the timbre that is expected of it. However, while this process may be hidden from both singer and audience, the *preferred* sound emitted from a vocal body, altered in accordance with sociocultural preferences, seems to confirm racialized, gendered, or other similar categories. The nature of such feedback loops, wherein singers adapt to others' perceptions of them, lies in their seeming confirmation of racialized ideas of vocal timbre. By focusing on the production of timbre, rather than on the timbre itself, and by dealing with the material reality of the singer and singing, we may avoid such traps.

The Shattering of La Stilla's Image and Voice

In his discussion of Jules Verne's *The Carpathian Castle* (1893)*,* Michel Poizat (1992: 182-84) posits the lost voice as the idealized voice. The voice that is lost acquires a potential to be adored that is not possible during a singer's lifetime, or indeed with the singer physically present. What is it about this quality of lostness that makes a voice available for idealization? Does a lack of presence clear an imaginative space which the listener may then enter, a space into which she may project fantasies that enable idealization? Or is the lost voice simply the voice we all hear when we listen, its individual articulation detached from the singer's body, lost and therefore found again in our minds?

The character around which the story coalesces is the Italian prima donna La Stilla. Her two admirers, the ageing Baron Rodolphe de Gortz and the young Romanian Count Franz de Telek, both attend her performances nightly. She is finally so deeply disturbed by Baron de Gortz's "silent intrusion" (1992: 97) that she feels she has no option but to terminate her public appearances and marry her second admirer, the young de Telek. On the night of her farewell performance she finally sees de Gortz's terrifying face, and dies of fright.

Years later, on a fateful night when the two men meet again, de Telek discovers that de Gortz has shut himself away in his castle, spending all his time with La Stilla's image and voice. In a prophetic flight of fiction Verne imagines that de Gortz's assistant, a technological genius, managed to steal both image and voice at the last concert. The voice is transmitted by a box, and the image is projected onto a series of mirrors, appearing in three-dimensional form. That night, in the encounter with another observer, the mirrors shatter and the voice becomes quiet.

Zero-G's singers Lola and Leon are the images and voices prophesied by Verne. Disassembled by de Gortz's gaze and reassembled when projected onto his mirrors, La Stilla was animated by de Gortz's mind. The singers providing thousands of phonemes to create Lola and Leon were also assembled in an attempt to create a coherent voice based on a stereotype. La Stilla's image shattered under the eyes and ears of a second observer, de Telek – while, as we may recall from users' comments, Vocaloid's generic soul voices burst semiotically in encounters with users. The sound may embody the singer with whom we believe we are in love––as de Gortz and de Telek believed––but any voice we hear is an individual articulation that may shatter in an encounter with others.

Conclusion

In light of performance theory, I have taken issue with the premise that the body with which we sing is unmediated.[17] I have argued that unlike a fingerprint, which is inherent to a particular body, vocal timbre is the sound of habitual performance that has *shaped* the physical body. Vocal timbre is not the unmediated sound of an essential body. Instead, both body and timbre are shaped by unconscious and conscious training practices that function as repositories for cultural attitudes toward gender, class, race, and sexuality. I have investigated racialized vocal timbre in order to assess both the production of vocal timbre and the construction of individual articulations of meaning and affect through it. Thus this work has examined the ideologies implicated in and transmitted through the body.

Underlying this work is a single goal: I hope to enhance our understanding of the interlocking and consequential relationships between the body, the act of singing and the medium of vocal timbre. I wish to untangle the processes involved in the construction of meaning through vocal timbre, and to distill their individual elements. The course of study I have undertaken is grounded in my conviction that by uncovering and analyzing the mechanisms involved in the production, reception and naming of vocal timbre we will also identify, and thereby denaturalize, the devices used in the construction and maintenance of racial stereotypes.

Two generations after government-mandated segregation ended in the United States, music software such as Vocaloid still bears witness to an active 'sonic color line' which is still very much alive and

timbre is not essential, but is instead a performed sound. I have therefore proposed an analysis of the performativity of timbre—performance as 'material creation of timbre' and 'reception as performance'—in order to begin to map and unravel the assumed connections between vocal timbre and race.

"The somatic realization of race," in the words of Deborah Wong, "is one of the great performative, destructive accomplishments of any society" (2000:87). I suggest that by relocating the search for the meaning of the voice from "the sound itself" to physical production of the sound and the processes that take place between the sound and the listener, we may begin the work of decolonizing vocal timbre and begin to recover the singer's agency.

Notes

[1] Quoted in Baugh (2003: 155).

[2] Edwards (1999); Purnell et al. (1999); Massey et al. (2001).

[3] For example, see 1999. *Clifford v. Kentucky. 7 SW 3d 371. Supreme Court of Kentucky*. In 1999, the Kentucky Supreme Court ruled that a conviction was appropriately based solely on a police officer's identification of a suspect whose voice the officer heard on an audio transmission. The officer identified the suspect as a black male and testified that during his 13 years as a policeman he had had several conversations with black men and therefore was able to identify the voice of a black male. In his ruling, the judge deduced that no one would find it inappropriate for an officer to identify the voice of a woman, and hence, "we perceive no reason why a witness could likewise identify a voice as being that of a particular race or nationality, so long as the witness is personally familiar with the general characteristics, accents or speech patterns of the race or nationality in question."
It may be argued that the police officer in this example relied mainly on accent, as opposed to timbre, which is the focus of this study. However, firstly, vowel variations is the basis for timbre and accents with its vowel specificities is therefore tied in with the resulting timbre; and, secondly, I use this example not to make a point in regards to vocal timbre, in specific, but to illustrate the belief that the voice (from language to accent and timbre) is intimately tied to the essential identity of the body of the speaker or singer in the given society.

[4] The first NAMM show was held on January 13-14, 2004. It has evolved to become one of the major international events introducing new music products. Vocaloid was first introduced through a demo in Musikmesse in Frankfurt in March 2003, and through Zero-G's website launched on October 23, 2003.

[5] Werde (2003).

[6] See http://www.vocaloid.com/en/introduction.html.

[7] Because of the translation that must take place between the written representation of a language and the sounded version of the language, the applications are language-specific. At this time the Vocaloid synthesis method is used only with English and Japanese; Lola, Leon and Miriam were programmed in English.

[8] In the terms outlined in footnote ix Vocaloid is not true vocal synthesis.

[9] In the late 1950s Bell Labs produced several speech synthesis systems which were capable of 'singing.' One of these systems, created by Kelly and Lochbaum in 1962, although too computationally intensive to be realistic for commercial use as speech synthesizer, was used in a collaboration with Max Mathews to generate early examples of singing synthesis (Wergo 1995). From this period of early speech signal processing, the channel vocoder (VOice CODER) and linear predictive coding (LPC) (Atal 1970; Makhoul 1975) were created. LPC created a revolution in speech synthesis and compositional possibilities. Some of its success was due to the similarity between the source/filter composition produced by the mathematics of linear prediction and the source/filter model of the human vocal tract. In the 1980s frequency modulation (FM) synthesis and formant wave function synthesis (FOF) were used for singing synthesis. FOF was later dubbed CHANT. In general terms, vocal synthesis may be divided into two different models, spectral and physical. The spectral model is, roughly speaking, based on perceptual mechanisms, and attempts to recreate the sound of the voice, while the physical model is based on production mechanisms and attempts to recreate the function of the voice (and, as a result, the sound). For more detailed information about different vocal synthesis models see Cook (1996).

[10] Practically, what matters to amateur users who neither know nor care about these distinctions, and to a general public told that the voice they hear is a synthesized voice, is not the technical distinction between full and hybrid vocal synthesis. What matters is that they *believe* it is vocal synthesis.

[11] A Japanese company, Crypton Future Media, Inc., released VOCALOID Meiko on October 5, 2004 based on the same synthesis method as the voices discussed in this paper. Zero-G released Vocaloid PRIMA, "a brand-new plug-in VIRTUAL VOCALIST modeled on the voice of a professional soprano opera singer, and powered by the all-new Yamaha VOCALOID 2 Singing Synthesis Technology" in January of 2007.

[12] Miriam Stockley is originally from South Africa, but moved to the United Kingdom in her teens to pursue a music career. In 1995 Stockley entered the spotlight with the album *Audiemus* in which her voice is recorded layer upon layer, producing a mix that is supposed to sound like "African voices."

[13] See http://www.vocaloid.com/en/index.html.

[14] For a critique of the highly problematic images used, please see Eidsheim (2008:109-117).

[15] To see a discussion of the commercial use of this piece, see Taylor (2000).

[16] Vocaloid's depictions of each of their synthetic voices feature very strong, offensive racial references. For a critique of this aspect of the software, please see Eidsheim (2008: 110-117)

[17] While vocal timbre is mediated both in performance and by the way we our listening organizes it, I do in no way reject the notion that in the sound of the voice there is also the expression of the uniqueness and singularity of a human being. Adriana Cavarero (2005) has written with deep insight about this subject.

Bibliography

Atal, B. 1970. "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave." *Journal of the Acoustical*

Americas, ed. S. Makoni, 155-168. London ; New York: Routledge.

Cavarero, Adriana. 2005. *For More than One Voice : Toward a Philosophy of Vocal Expression*. Stanford, Calif.: Stanford University Press.

Cook, Perry R. 1996. Singing Voice Synthesis: History, Current Work and Future Directions. *Computer Music Journal* 20 (3):38-46.

Clifford v. Kentucky. 1999. *7 SW 3d 371. Supreme Court of Kentucky.*

Edwards, John. 1999. "Refining our Understanding of Language Attitudes." *Journal of Language and Social Psychology* 18 (1):101-110.

Eidsheim, Nina Sun. 2008. "Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance." Dissertation, Music, University of California, San Diego, San Diego.

Hall, Stuart. 1980. "Race, Articulation and Societies Structured in Dominance." In *Sociological Theories: Race and Colonialism*, ed. Unesco, 305-345. Paris: Unesco.

Hall, Stuart and Lawrence Grossberg. 1986. On Postmodernism and Articulation: An Interview with Stuart Hall." *Journal of Communication Inquiry* 10: 54.

Keefe, Dom. 2005-2006. Email exchanges.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things : What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Leland, Elizabeth. 2002. Born White, Raised Black: Linda McCord Always Believed She Was African-American - Until the Phone Rang One Day. *The Charlotte North Carolina Observer Newspaper*, June 30, 2002, 1G.

Makhoul, J. 1975. "Linear Prediction: A Tutorial Review." *Proceedings of the IEEE*. 63: 561-580. New York: Springer.

Massey, Douglas and Garvey Lundy 2001. "Use of Black English and Racial Discrimination in Urban Housing Markets: New Methods and Findings." *Urban Affairs Review* (36):470-96.

Middleton, Richard. 1990. *Studying Popular Music*. Milton Keynes [England] ; Philadelphia: Open University Press.

———. 2003. Locating the People: Music and the Popular In *The Cultural Study of Music : A Critical Introduction*, edited by M. Clayton, T. Herbert and R. Middleton. New York: Routledge.

———. 2006. *Voicing the Popular : On The Subjects of Popular Music*. New York: Routledge.

Niedzielski, Nancy. 1999. The Effects of Social Information on the Perception on Sociolinguistic Variable. *Journal of Language and Social Psychology* 18 (1):62-85.

Poizat, Michel. 1992. *The Angel's Cry : Beyond the Pleasure Principle in Opera*. Ithaca: Cornell University Press.

Purnell, Thomas, William Idsardi, John Baugh. 1999. "Perceptual and Phonetic Experiments on American English Dialect Identification." *Journal of Language and Social Psychology* 18 (1):10-30.

Rubin, D.L. 1992. Nonlanguage Factors Affecting Undergraduates' Judgments of Nonnative English-speaking Teaching Assistants. *Research in Higher.*

Shepherd, John, and Peter Wicke. 1997. *Music and Cultural Theory*. Cambridge England Malden, Mass.: Polity Press ; Published in the USA by Blackwell Publishers.

Sodergren, Anders 2006. Vocaloid. Edinburgh, September 28, 2006.

Tagg, Philip. "Vocal Persona." In *Music's Meaning* (book in progress, unpublished). Accessed June 10, 2008 from http://tagg.org/bookxtrax/NonMuso/NM09.pdf

Taylor, Timothy Dean. 2000. World Music in Television Ads. *American Music* 18 (2):162-192.

Vocaloid:User. http://www.vocaloid-user.net/ (Accessed June 25, 2006).

Werde, Bill. 2003. "Could I Get That Song in Elvis, Please?" *New York Times Online*, November, 23.

Wergo. 1995. *The Historical CD of Digital Sound Synthesis.* Schallplatten GmbH, Mainz, Germany. WER 2033-2.

Wong, Deborah Anne. 2004. *Speak it Louder : Asian Americans Making Music*. New York: Routledge.

subir >

x?