Christine Bauer, Johanna Devaney

Constructing Gender in Audio: Exploring how the curation of the voice in music and speech influences our conception of gender identity

Abstract: In diesem Artikel untersuchen wir die explizite Vergeschlechtlichung in der Art und Weise, wie Stimmen im Kontext von Musik und Audio behandelt werden, und analysieren, inwiefern dies mit der speziellen Funktion der Stimme in einem gegebenen Kontext zusammenhängt. Aufbauend auf bestehenden Arbeiten zu Gender in der Singstimme untersuchen wir, wie die Stimme durch den Einsatz von Software zur Stimmproduktion („Voice Production Software") genderspezifisch geprägt wird. Insbesondere betrachten wir Auto-Tune, das die Bearbeitung einer aufgenommenen echten Stimme erlaubt, und Vocaloid, das es ermöglicht, computergestützt eine neue Singstimme zu erzeugen. Des Weiteren untersuchen wir Parallelen zur Sprachtechnologie bei interaktiven Sprachassistenten mit künstlicher Intelligenz. Diese Sprachtechnologie weist insofern Parallelen zu den oben genannten Musiktechnologien auf, als es sich auch hier entweder um bearbeitete Aufnahmen einer natürlichen Stimme handelt oder um eine computergestützt neu erzeugte Stimme. Unsere Analyse zielt darauf ab, über die binäre Betrachtung von Gender hinauszugehen und die intersektionalen Identitäten von Stimmen in den Bereichen Musik und Sprache zu berücksichtigen. Diese Analyse der Kuratierung von Stimmen in Musik und Sprache trägt zum Verständnis der klanglichen Konstruktion von Gender in unserer Gesellschaft bei.

Abstract: In this article, we explore explicit gendering in the manner in which voices are treated in music and audio and whether this relates to the specific function of the voice in a given context. Building on existing work on gender in singing, we explore the ways in which the voice is gendered through the use of voice production software. Specifically, we look at Auto-Tune, which allows for a recorded natural voice to be manipulated, and Vocaloid, which allows for the computational generation of new singing voices. We also examine parallels in speech in terms of interactive artificial intelligence voice assistants. This speech technology parallels the aforementioned music technologies in that the voice may be either manipulated recordings of natural voice or computational generations. Our analysis aims to look beyond gender binaries and to consider the intersectional identities of the voices

in the fields of music and speech. This analysis of the curation of voices in music and speech contributes to our understanding of the aural construction of gender in society.

1 Introduction

The development of recording technology in the late 19th century allowed for voices to be heard separate from their physical body and the physical space in which the recordings were made. As technology progressed, techniques for manipulating and augmenting recording signals were developed that allowed for voices to become further disembodied. This article considers the way in which recent technologies that manipulate recorded voice in music and speech have gendered implications. Specifically, we will consider how notions of gender are constructed through a combination of the aural presentation of the voice and other factors, including visual and functional factors related to the representation and delivery of the voice.

In music, we will consider the role of technology that can manipulate vocal pitch and timbre (such as Auto-Tune) and computationally generate singing voices (such as Vocaloid) in the construction of gender in singing. With pitch and timbral manipulation technology, we will examine the voice qualities that are created with this technology and how both the creation of these voices and their reception differs by gender. With synthesis technology, we will consider how gender is largely conferred by the representations and identities associated with the audio rather than the acoustics of the audio itself. This is paralleled in speech technology, where we will examine the increasingly pervasive artificial intelligence (AI) voice assistant interfaces and how the marketing and function of these devices define the gender of these interactive assistants beyond the acoustics of the generated voices.

We will also look beyond a gender binary. We will describe how particularly the timbral aspects of this technology have been used both as a supportive means to achieve stereotypical gender effects and also in a creative manner to play with and, ultimately, break gender stereotypes. Voice technology has also been applied to create a presumptively "genderless" voice, and we will examine how such a gender-ambiguous voice is perceived and show that various cues within and beyond a voice create a gendered "body" in the listener's mind.

We will begin this article with a brief description of what the term gender connotes (Section 2) and give an overview of the history of voice technology (Section 3). Delving into detail, we will examine gender in pitch-manipulation technology

(Section 4), voice assistants (Section 5), and synthetic singing voices (Section 6). We will dedicate Section 7 to discussing voice technology specifically beyond the gender binary. Finally, we will conclude this work by reflecting on the ways in which the interplay of voice and gender in the technologies we have considered connects to other technologies.

2 What is Gender?

When talking about gender, it is important to clarify the two constructs, sex and gender. Colloquially they are often used interchangeably or understood as referring to one and the same (Kessler & McKenna 1978); however, formal discourse distinguishes between these constructs. Sex is understood as a biological category that is derived from a person's anatomy and is traditionally assigned at the time of birth[1] (C. West & Zimmerman 1987) whereas the term gender is typically used to refer to a person's behavior and social role (Keyes 2018).

There is a traditional view – and according to Burtscher & Spiel (2020), it is still considered the most prevalent view on gender within society – that a person's sex is considered a binary category with the two options "male" and "female"; that a person's gender is an inevitable consequence of their sex (C. West & Zimmerman 1987; Skewes, Fine, & Haslam 2018). The binary conception of sex leads to a binary conception of gender, with the two options "man" and "woman" (Schilt & Westbrook 2009). The essentialist view considers a person's sex/gender attribution as being inevitably fixed (Bohan 1993). A more contemporary view considers gender as socially constructed and performed, where the sexed body determines how gender is supposed to be performed (Nicholson 1994); yet again a binary view is assumed – that people identify with one of two genders. Research has repeatedly shown, however, that this binary view is inaccurate (Keyes 2018; Messerschmidt 2009; Hyde et al. 2019; Bornstein 2016; Haynes & McKenna 2001).[2] Thus, gender theorists have recognized that gender (and sex) go far beyond a simple binary and understand it as a spectrum rather than a binary. A recent overview of the non-binary view of gender by Fausto-Sterling (2020) shows this view has only recently been acknowledged on a wider basis. Thereby, the term non-binary is used as an umbrella term to refer to identities that fall outside of or between the male and female identities drawn from the binary conception of gender, and transitioning from one gender to another is not necessarily strictly female-to-male or male-to-female (Matsuno & Budge 2017; Monro 2019).

Against this background, we want to emphasize that this article does not intend to argue that gender is binary. On the contrary, we describe how literature reflects how gender is constructed and transformed with voice technology – and how gender is perceived. We note that literature on voice technology frequently adopts a binary conception when discussing gendered voices. While the binary view is consistent with overriding structures in society (which have only recently begun to acknowledge non-binary identities) and may also relate to how voices are currently perceived, we aim to problematize this binary concept and show how it limits our understanding of the relationship between voice and gender.

3 A Brief History of Voice Technology

Magnetic tape was initially developed in the 1920s and 1930s for simple sound reproduction. With the development of the Magnetophon in the 1940s, art-music composers used magnetic tape to generate new sounds by splicing pieces of tape together and manipulating its playback. These techniques were subsequently adopted by commercial recording studios in the 1950s and 1960s, where multiple recorded performances were spliced together, substituting perceived errors in one performance with better renditions in another. This process, however, was laborious, requiring tape to be accurately cut into small segments from the longer segments of audio, sometimes into segments as small as individual notes.

Technologies for analyzing existing voices and synthesizing new ones were also developed in the first half of the 1900s, and could be applied not only to generate new voices but also to manipulate the sound of the voice. The vocoder is an example of the latter. It was developed in the 1930s and widely used in World War II for safe transmission of speech across telecommunication platforms. The vocoder (which is a blend of the words voice and encoder) is a voice codec for speech compression that analyzes natural speech, encodes it, and resynthesizes speech at the recipient. The vocoder process introduced certain alterations to the pitch and timbre of the original sound, creating a distinct, often robotic sound that was adopted by musicians for artistic means in the 1970s (Tompkins 2011). Around the same time, more sophisticated analysis/synthesis technology was developed, such as Eventide's H910 harmonizer device, which was introduced in 1974 and allowed for semitone retuning of the voice within the range of an octave. Similarly, the advent of the sampler in the 1970s added greater flexibility as individual notes could be recorded, manipulated, and played back. These effects and sampler devices, however, often distorted the timbre of the voice, creating slightly unnatural

timbres. Thus, it was audible when such effects and samples were used, which was typically intentional.

In the 1980s, more sophisticated digital signal algorithms were developed that could shift pitch while better maintaining the timbre of the voice than earlier technologies. Examples include the pitch-synchronous-overlap-and-add (PSOLA) algorithm that was first developed in the 1980s for speech synthesis (Charpentier & Stella 1986) and subsequently applied to voice pitch-shifting the following decade (Moulines & Laroche 1995). As with the vocoder, many of these technologies were initially developed for speech technology but subsequently applied to singing and, often, musical instruments. Auto-Tune was released in 1997 by Antares, building on auto-correlation for pitch estimation (Rabiner 1977) – a technology that was developed for speech during the 1970s. Auto-Tune allowed users to retune vocals much more precisely than previously available technology, down to one-hundredth of a semitone. It also allows for users to control how fast the pitch was tuned after the start of the note. Since natural voices tend to take some time to settle on the pitch at the start of each note, setting this parameter to zero resulted in a robotic voice that sounded like a vocoder.[3] Subsequently, several other pitch-correction software tools were released, most notably Celemony's Melodyne and Waves' Tune. Melodyne, in particular, has become an industry-standard because of its ability to correct pitch with less timbral distortion than Auto-Tune; in other words, it can correct pitch more invisibly.

Synthetic voice technology has a similar history to analysis/synthesis technology in that here, too, research and development were initially focused on speech. The early goal of this technology was to computationally generate understandable speech, starting in the 1930s with Bell Laboratory's human-controlled source-filter Voder speech synthesizer and continuing with the development of speech synthesizers that attempted to model the vocal tract in the 1950s (see Klatt 1987 for a summary of the early history of speech synthesis). Recent developments have focused on expressive speech, creating a synthetic speech that sounds more natural (see Schröder 2009 for a summary of early developments in expressive speech). In the past decade, this technology has been used extensively in virtual AI assistants, such as Amazon's Alexa, Apple's Siri, and Cortana by Microsoft, as well as the Google Home device.

Researchers began working on singing synthesis in the 1970s, with a large amount of the work being undertaken in Sweden at KTH (Larsson 1977) and in Paris at IRCAM (Rodet, Potard, & Barriere 1984). But given the complexities of generating a rich and natural synthetic singing voice, not much headway was made in terms of

commercial products until the 2000s. A breakthrough in singing generation came with Yamaha's Vocaloid technology, which was based on an analysis/synthesis approach to generating new singing voices from existing voices. The technology was developed by Bonada and colleagues at the Universitat Pompeu Fabra (Bonada et al. 2001) and commercially licensed and produced by Yamaha. Vocaloid was the first commercial product that offered musicians access to a fully synthesized singing voice (Bell 2016). Vocaloid consists of two components: the commercially developed singing synthesis software by Yamaha (i.e., Vocaloid software) and voice libraries, which are typically developed and released by third-party companies (Kenmochi 2010). A user inputs melody and text information into the Vocaloid software to then be auralized with the selected Vocaloid voice. While Vocaloid is typically described as a completely artificially created singing voice, its basis in the analysis/synthesis system by Bonada et al. (2001) means that each Vocaloid voice is in fact seeded by recordings of a natural human voice. A Vocaloid voice contains samples of all possible vowel and consonant combinations – in the English language, there are about 3,800 possible vowel and consonant combinations recorded by a single person (Eidsheim 2009).

4 Gender and Pitch-Manipulation Technology

The first major commercial usage of Auto-Tune for pitch-correction also showcased the timbral distortions that can arise with extreme software settings. The dance-track *Believe* by Cher was released in 1998 to great commercial success alongside mixed reviews from critics (e.g., Brackett & Hoard 2004). Starting shortly after its release, *Believe* has been subject to a great deal of academic analysis. Early examination of Cher's vocals conflated Auto-Tune's timbral distortion with effects attainable with a vocoder. In music, the vocoder had been used throughout the 1970s, 1980s, and 1990s by predominantly male vocalists, such as Kraftwerk and Afrika Bambaataa, to produce a robotic-sounding voice (Dickinson 2001; Heesch 2016), the main exceptions being the use of the vocoder by avant-garde composers Wendy Carlos and Laurie Anderson. Dickinson (2001) explored the relationship between the excesses of the cyborg-like voice (created by both the standard vocoder and the vocoding-sounding autotune[4] effect) and camp[5] in gay culture. Specifically, she juxtaposed the almost exclusive use of the vocoder by straight males in popular music with Cher's diva persona. Heesch (2016) considered how the early male domination of vocoder-like effects was linked to how both female and African-American singing voices have traditionally been viewed as more natural, specifically rooted in the body, than those of white males. The link between women and naturalness was also examined by Cadilhe (2016), who explored the relationship between autotuning and a carnival

tradition. Cadilhe (2016) considered Cher's campy performance to be an example of gender parody and argued that she was playing with the traditional association of men with science and women with naturalness.

A number of scholars have explored the difference between the use of Auto-Tune to create artificial timbral effects and the invisible use of Auto-Tune to correct performances. The "invisible practice" of pitch correction as described by Marshall (2018) is much more prevalent with female voices, arguably because it results in a more natural voice than an application of Auto-Tune that results in timbral distortions. Provenzano (2019a, 2019b) argues that the invisible use of pitch correction to create perfect, yet natural sounding, female voices stands in direct contrast to the "artistic" or "creative" use of Auto-Tune by male artists such as T-Pain, who used the software to create a distinct robot-sound voice. Indeed, although using Auto-Tune to correct pitch can be considered an open secret in the music industry, its specific use for specific artists is rarely discussed. There is also the (artistic) pressure that female voices are expected to be perfect, and the manner in which critique derides female singers whose voices have been pitch-corrected. Coulter (2017) studied the views of pre-teenage girls on Auto-Tune and found that they were generally negative about the technology, believing a natural (and ideally perfect) voice to be better. These negative associations with pitch correction were also seen in television's X-Factor Auto-Tune controversy in 2011, where the public backlash to the apparent autotuning applied to female contestant Gamu Nhengu's performance led to the producers promising to discontinue its use in the future.

An important question that arises is how much agency female singers have in the autotuning of their voices. Provenzano (2019a, 2019b) examined how the male domination of the recording studios creates an environment where the female voice is curated by predominantly male recording engineers. In her ethnography of record producers in Los Angeles, she found widespread instances of non-consensual re-tuning of female vocalists by male engineers. Even when consent is given, it is reasonable to assume that is not given completely freely. The now-standard application of Auto-Tune to female voices has created an expectation of a perfect voice; female singers who do not fulfil this expectation are often regarded as sounding not "good" or not (sufficiently) female, which leads to the further use of Auto-Tune, consensually or non-consensually, reproducing more of the same (i.e., perfect autotuned voice for females).

5 Gender and AI Voice Assistants

The notion of the "perfect" female voice is relevant to the predominant use of female-sounding voices in AI voice assistants. A recent report by UNESCO (M. West, Kraut, & Chew 2019) examined the assumed female identities of the major technologies in the AI voice assistant market: Alexa by Amazon, Cortana by Microsoft, and Siri by Apple. The report describes not only the feminized names and sounds of these assistants but also their feminized backstories and "helpful" nature, even as the tech companies claim that the devices are genderless. These observations are supported by the work of Hannon (2016), who has examined how the words used by female AI voice assistants, particularly the extended use of first-person phrasing, creates a sense of subordinance that is reinforced by (and likely also reinforcing) their perceived gender. Obinali (2019) takes a similar view and observes the similarities between the AI voice assistant's role and that of a secretary, a role traditionally held by women. The development of these AI virtual assistants by primarily male researchers and engineers[6] parallels male dominance in music recording studios. In both cases, the female voice is crafted predominantly by men – with limited (in the case of singers) or no (in the case of AI virtual assistants) input from the resultant feminized voice.

6 Gender and Synthetic Singing Voices

Since its release in 2004, Vocaloid has been the dominant commercial tool for generating synthetic singing. As noted above, the core Vocaloid product is a synthesis engine with a wide range of voice libraries that can be purchased separately. Each Vocaloid voice library is typically marketed with an assigned profile or ascribed attributes. This may range from more general categorizations in terms of music genre or the voice's gender, up to a detailed personal profile of the fictional character behind the voice (Eidsheim 2009; Roseboro 2019). These descriptions shift the disembodied Vocaloid voice to embodied Vocaloid character. The most popular Vocaloid character is Hatsune Miku, which was released in August 2007 (Kenmochi 2010). It was the first Vocaloid voice library that, apart from voice samples, featured a cute 3D anime-style character. Since then, Hatsune Miku has become the most prominent representation of the Vocaloid phenomenon (Kenmochi 2010; Roseboro 2019) with fans using the Vocaloid character's library to create new original content on a large-scale basis (Kenmochi 2010; Klein 2016), peaking in hologram concerts with the character (Roseboro 2019). Sabo (2019) argues that while Hatsune Miku's voice corresponds to certain gender norms we perceive as female or feminine, her

gender in relation to her embodied appearance is shaped as much by Japanese society and culture as by the acoustics of her voice.

Vocaloid has a large number of parameters for adding vocal effects such as vibrato, pitch glide, and timbre (Jude 2018), and these parameters can be adjusted to manipulate the sound and characteristics of the voice continuously throughout a composition (Bell 2016). Several of these parameter settings have presets with verbal descriptions. For example, the timbral option "vivid" is described as a "bright, cheerful voice" while the timbre "light" is described as an "innocent, heavenly voice" (Jude 2018). One of the editable parameters is called "gender factor". Adjusting this parameter applies a combination of filters that affect pitch transposition, timbre mapping, and spectral shape (Bell 2016). The product catalogue[7] offers more female Vocaloid voices than male voices, and some voices come in male-female duos of separate voice libraries. Interestingly, many of the gender-paired duos are samples from a single voice actor using different vocal registers. For instance, Shimoda Asami, whose voice was sampled for the Vocaloid twins Kagamine Rin and Len, explained that when recording the samples for the male voice, Len, she spoke from her belly, whereas she spoke at the top of her head for recording the female voice, Rin (Bell 2016). Thus, while a Vocaloid voice library is constructed of sound samples that are gendered by the identity of the singer being sampled (such as Shimoda Asami), the gender of the Vocaloid (such as Len) is largely conferred by the visual representation, or embodiment, of the voice rather than the acoustics. A Vocaloid voice, which is disembodied by nature, produces a "body" in the listener's mind (Connor 2000; Ferrete-Vázquez 2019), where, as Eidsheim (2009) argues, a body that is produced in such a way is always embedded in and framed by the listener's concepts. Gender is not inherent in a voice; in embodied voices, visual (e.g., hairstyle and hair length, make-up, accessories, gesture, body posture) and behavioral factors (e.g., how long someone talks for or how often they speak) contribute to the gender assessment (Sutton 2020). Some factors are more influential than others, although to date, the hierarchy of these factors is not well understood (Nass & Lee 2000; Sutton 2020). Young (2012) further observed that it is also common for Vocaloid users to play with voices so that these are ambiguous as to gender, with voices more appropriately placed along a continuum between male and female, rather than as one or the other. This raises the question of the degree to which we assign a binarized gender to a voice, even when acoustically the gendering is ambiguous.

7 Beyond a Gender Binary

Jude (2018) argued that we learn to associate certain vocal sounds (including pitch and timbre, but also linguistic and paralinguistic elements) with gendered bodies; and building on these unconscious associations, we build a construct of what gendered bodies must sound like. Thereby, the notions that we (unconsciously) observe and the associations we draw are dependent on and shaped by our social and cultural environment and interactions. Eidsheim (2009), for instance, postulates that vocal timbre is an artefact of identity – a performance – rather than biology, and Ferrete-Vázquez (2019) suggests that we learn social markers of voices (e.g., gender, ethnicity) and assign those markers to the voices we hear. For example, vocal formants suggest the internal size and shape of the vocal tract, which relates to body size and thus implies a gender, based on men typically being larger than women, but, as Jude (2018) observes, these do not determine rigidly binary-gendered bodies.

A voice also transmits gender cues by the language used (e.g., word choice), the topic or the message conveyed (Jude 2018). Sutton (2020) examined 'Q The First Genderless Voice' in relation to the aforementioned UNESCO report on gender in AI voice assistants (M. West, Kraut, & Chew 2019), observing that even if markers of gender are removed from the acoustic voice that there are still gendered elements in the design. In AI voice assistants, this is reinforced by gendered markers in the speech patterns and the function of the assistive technology to perform tasks typically performed by females (Costa & Ribas 2019). Thus, attempts to break beyond the prevalent binary gender conception is challenging, as this binary is deeply rooted in large parts of society. Yates (2020) points out how Q's voice was evaluated: Participants were asked what they perceived to be Q's (binary) gender and the answers were split 50/50. So while the gender perception was ambiguous on average, on the level of the individual, every respondent perceived a specific gender in Q; they perceived either a female or a male voice. Thus, it appears that a gender-ambiguous voice is – still – absorbed into the binary conceptualization of gender, where it can be "drawn in" into one category or the other (Sutton 2020).

Considering that a majority of human bodies – irrespective of their gender – are able to create a mid-range of pitch and may learn to control articulation to achieve a 'transgender effect' (Bell 2016) or what may be called a 'gender-ambiguous voice', the question remains when we will acknowledge and absorb the wide range of gender identities in our perceptions. This effect has been explicitly explored by several trans singers through their use of vocal manipulation technology. Blanchard (2018) discusses SOPHIE, the late Scottish trans femme electronic artist and singer

who made use of pitch-shifting technology to raise her vocal pitch, and compares this to Snapchat filters that achieve similar voice heightening effects.[8] Gratton (2016) describes the conscious identity work of non-binary identities in order to avoid being misgendered as one's gender-assigned-at-birth. In non-queer spaces, using and exaggerating (binary) gender stereotypes – also in linguistics – is a means to position themselves in relation to gender binaries. Blanchard (2018) notes that Snapchat audio filters are sometimes used similar to hormone therapies in that they are applied to give the person transitioning a more stereotypical-sounding voice for the gender that they are transitioning to. She also states that vocal manipulation technologies are used creatively by trans artists to play with preconceived acoustic notions related to the binarized conception of gender.

8 Conclusion

Voice technology, like all technology, is made and consumed by humans. Thus, technology is never neutral and is rooted in preconceptions that the developers and users bring to it. In the case of voice and gender, this both influences how the acoustics of the voice are generated and manipulated, and how the resultant voices are perceived. There is also a feedback loop that emerges between creation and reception, with the way in which technology is used influencing further development of the technology. In the case of autotuning, the desire for more natural-sounding, invisible corrections has led to the further development of both Auto-Tune and competing software (such as Celemony's Melodyne and Wave's Tune), which allow music producers greater control over the pitch manipulation through more sophisticated algorithms and graphical user interfaces that allow for more precise control. There is also a feedback loop in between what we hear and what we vocally produce, both in the recorded and natural worlds (Eidsheim 2009). In the case of singing technology, this can lead to expectations of perfection in both recordings and natural performances, influencing both real and synthetic singing. In the case of AI voice assistants, this can lead to the reinforcement of the traditionally subservient role of more feminized people as AI voice assistants perpetuate the association of the female voice with an assistive role. In all cases, it can be observed that it is predominantly male engineers who develop these technologies, and this has an outsized impact on how female voices are manipulated, generated, and disseminated. This results in the perpetuation of a specific gendering, both in terms of what is considered female and in terms of the adherence to strict gender binaries.

The issues of gender-imbalances in the development and dissemination of technology is not limited to voice technology. For example, Vásárhelyi (2020) has discussed the

impact of gender imbalances in development teams for video games and how these impact the final product. Nor are the imbalances limited to gender identities, as has been widely noted in work on Ethical AI, such as Gebru's (2020) discussion of issues in AI facial recognition at the intersection of gender and ethnicity, and the examination by Whittaker et al. (2019) of how concepts of "normal" and "ability" are encoded into AI systems and how this may impact people with disabilities. Dillon & Collett (2019) proposed that a four-part approach is necessary to address these intersectional issues. Their approach includes not only improving gender-balance in the AI workforce and bias in AI datasets but also incorporating gender theory more explicitly and looking beyond the technology itself to how AI is governed in society through law and policy.

In sum, the topic of voice technologies and gender is less about how computers are used to replicate stereotypical markers of gender (or ethnicity) in natural voices, and more about how the output of the technologies are consciously or unconsciously applied to present voices in a manner that influences how listeners perceive a certain gender (or ethnicity). Thereby technology does not only refer to voice technologies in isolation but embraces all sorts of technologies and mechanisms (e.g., visual manifestations, type of tasks performed, and word choice) that embody the voice. To understand gendered perception and to break the feedback loop, we see two critical paths to follow. First, more research is needed to understand the hierarchy of cues (acoustic, visual, etc.) that shape gender perception. Second, we need diverse groups of people researching and developing voice technologies – and we need them to apply these technologies in the field.

References

Bell, Sarah A. (2016): The dB in the .db: Vocaloid Software As Posthuman Instrument. In: Popular Music and Society. 39/2. Pp. 222-240. DOI: 10.1080/03007766.2015.1049041.

Blanchard, Sessi K. (2018): How SOPHIE and Other Trans Musicians Are Using Vocal Modulation to Explore Gender. Pitchfork.com. June 28. https://pitchfork.com/thepitch/how-sophie-and-other-trans-musicians-are-using-vocal-modulation-to-explore-gender/ [last accessed July 6, 2021].

Bohan, Janis S. (1993): Regarding Gender: Essentialism, Constructionism, and Feminist Psychology. In: Psychology of Women Quarterly. 17/1. Pp. 5-21. DOI: 10.1111/j.1471-6402.1993.tb00673.x.

Bonada, Jordi; Celma Herrada, Òscar; Loscos, Àlex; Ortolà, Jaume; Serra, Xavier; Yoshioka, Yasuo; Kayama, Hiraku; Hisaminato, Yuji; Kenmochi, Hideki (2001): Singing Voice

Synthesis Combining Excitation Plus Resonance and Sinusoidal Plus Residual Models. In: Proceedings of the 2001 International Computer Music Conference (ICMC 2001). Havana, Cuba. Michigan: Michigan Publishing.

Bornstein, Kate (2016): Gender Outlaw: On Men, Women, and the Rest of Us. 2nd revised and updated edition. New York, NY, USA: Vintage.

Brackett, Nathan; Hoard, Christian D. (2004): The New Rolling Stone Album Guide. 4th edition. London, UK: Simon & Schuster.

Bridges, Chandler R., Jr. (2020): Effects of Software Tuning Programs on Vocal Recordings. In: Proceedings of the Audio Engineering Society Convention 149 (AES 149). Paper No. 10436.

Burtscher, Sabrina; Spiel, Katta (2020): "But Where Would I Even Start?": Developing (Gender) Sensitivity in HCI Research and Practice. In: Proceedings of the Conference on Mensch und Computer (MuC 2020). Magdeburg, Germany. Pp. 431-441. DOI: 10.1145/3404983.3405510.

Cadilhe, Orquídea (2016): Cher's Music Videos. Gender As a Performative Construction. In: Gender in Focus: (New) Trends in Media. Carla Preciosa Braga Cerqueira; Rosa Cabecinhas; Sara Isabel Magalhães. (eds.). Braga, Portugal: Universidade do Minho. Centro de Estudos de Comunicação e Sociedade (CECS). Pp. 103-121.

Charpentier, Francis; Stella, M.G. (1986): Diphone Synthesis Using an Overlap-Add Technique for Speech Waveforms Concatenation. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (CASSP '86). Tokyo, Japan. Pp. 2015-2018. DOI: 10.1109/ICASSP.1986.1168657.

Connor, Steven (2000): Dumbstruck: A Cultural History of Ventriloquism. New York, NY, USA: Oxford University Press.

Costa, Pedro; Ribas, Luísa (2019): AI Becomes Her: Discussing Gender and Artificial Intelligence. In: Technoetic Arts. 17/1-2. Pp. 171-193.

Coulter, Bridget (2017): Singing from the Heart: Notions of Gendered Authenticity in Pop Music. In: The Routledge Research Companion to Popular Music and Gender. Stan Hawkins (ed.). New York, NY, USA: Routledge. Pp. 285-298.

Dickinson, Kay (2001): 'Believe'? Vocoders, Digitalised Female Identity and Camp. In: Popular Music. 20/3. Pp. 333-347.

Dillon, Sarah; Collett, Clementine (2019): AI and Gender: Four Proposals for Future Research. https://www.repository.cam.ac.uk/handle/1810/294360 [last accessed July 12, 2021]. DOI: 10.17863/CAM.41459.

Eidsheim, Nina S. (2009): Synthesizing Race: Towards an Analysis of the Performativity of Vocal Timbre. In: Trans. Revista Transcultural de Música. 13. Pp. 1-9.

Fausto-Sterling, Anne (2020): Sexing the Body: Gender Politics and the Construction of Sexuality. 2nd updated edition. New York, NY, USA: Basic Books.

Ferrete-Vázquez, Jaume (2019): Bodies Reappear As Action: On Synthetic Voices in Performance. In: Performance Research. 24/7. Pp. 123-129. DOI: 10.1080/13528165.2019.1717880.

Gebru, Timnit (2020): Race and Gender. In: The Oxford Handbook of Ethics of AI. Markus D. Dubber; Frank Pasquale; Sunit Das (eds.). New York: Oxford University Press. Pp. 251-269. DOI: 10.1093/oxfordhb/9780190067397.013.16.

Gratton, Chantal (2016): Resisting the Gender Binary: The Use of (ING) in the Construction of Non-binary Transgender Identities. In: University of Pennsylvania Working Papers in Linguistics. 22/2. Pp. 51-60. Paper No. 7.

Hannon, Charles (2016): Gender and Status in Voice User Interfaces. In: Interactions. 23/3. Pp. 34-37. DOI: 10.1145/2897939.

Haynes, Felicity; McKenna, Tarquam (eds.) (2001): Unseen Genders: Beyond the Binaries. New York, NY, USA: Peter Lang Publishing.

Heesch, Florian (2016): Voicing the Technological Body: Some Musicological Reflections on Combinations of Voice and Technology in Popular Music. In: Journal for Religion, Film and Media. 2/1. Pp. 49-69. DOI: 10.25364/05.2:2016.1.5.

Hyde, Janet S.; Bigler, Rebecca S.; Joel, Daphna; Tate, Charlotte C.; van Anders, Sari M. (2019): The Future of Sex and Gender in Psychology: Five Challenges to the Gender Binary. In: American Psychologist. 74/2. Pp. 171-193. DOI: 10.1037/amp0000307.

Jude, Gretchen (2018): Vocal Processing in Transnational Music Performances, from Phonograph to Vocaloid. PhD thesis. University of California, Davis.

Kenmochi, Hideki (2010): VOCALOID and Hatsune Miku Phenomenon in Japan. In: Proceedings of the First Interdisciplinary Workshop on Singing Voice (InterSinging 2010). Tokyo, Japan. https://www.isca-speech.org/archive_v0/int_singing_2010/papers/isi0_001.pdf [last accessed July 12, 2021].

Kessler, Suzanne J.; McKenna, Wendy (1978): Gender: An Ethnomethodological Approach. New York, NY, USA: Wiley.

Keyes, Os (2018): The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. In: Proceedings of the ACM on Human-Computer Interaction. 2(CSCW). Pp. 1-22. Paper No. 88. DOI: 10.1145/3274357.

Klatt, Dennis H. (1987): Review of Text-to-Speech Conversion for English. In: The Journal of the Acoustical Society of America. 82/3. Pp. 737-793. DOI: 10.1121/1.395275.

Klein, Eve (2016): Feigning Humanity: Virtual Instruments, Simulation and Performativity. In: IASPM Journal. 6/2. Pp. 22-48.

Larsson, Bjorn (1977): Music and Singing Synthesis Equipment (MUSSE). In: Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR). 18/1. Pp. 38-40.

Marshall, Owen (2018): Auto-Tune in Situ: Digital Vocal Correction and Conversational Repair. In: Critical Approaches to the Production of Music and Sound. Samantha Bennett; Eliot Bates (eds.). New York, NY, USA: Bloomsbury Publishing. Pp. 175-194.

Matsuno, Emmie; Budge, Stephanie L. (2017): Non-binary/Genderqueer Identities: A Critical Review of the Literature. In: Current Sexual Health Reports. 9/3. Pp. 116-120. DOI: 10.1007/s11930-017-0111-8.

Messerschmidt, James W. (2009): "Doing Gender": The Impact and Future of a Salient Sociological Concept. In: Gender & Society. 23/1. Pp. 85-88. DOI: 10.1177/0891243208326253.

Monro, Surya (2019): Non-binary and Genderqueer: An Overview of the Field. In: International Journal of Transgenderism. 20/2-3. Pp. 126-131. DOI: 10.1080/15532739.2018.1538841.

Moulines, Eric; Laroche, Jean (1995): Non-Parametric Techniques for Pitch-Scale and Time-Scale Modification of Speech. In: Speech Communication. 16/2. Pp. 175-205. DOI: 10.1016/0167-6393(94)00054-e.

Nass, Clifford; Lee, Kwan M. (2000): Does Computer-Generated Speech Manifest Personality? An Experimental Test of Similarity-Attraction. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00). The Hague, The Netherlands. Pp. 329-336. DOI: 10.1145/332040.332452.

Nicholson, Linda (1994): Interpreting Gender. In: Signs: Journal of Women in Culture and Society. 20/1. Pp. 79-105. DOI: 10.1086/494955.

Obinali, Chidera (2019): The Perception of Gender in Voice Assistants. In: Proceedings of the Southern Association for Information Systems Conference (SAIS 2019). St. Simon's Island, GA, USA. Paper No. 39.

Provenzano, Catherine (2019a): Emotional Signals: Digital Tuning Software and the Meanings of Pop Music Voices. PhD thesis. New York University.

Provenzano, Catherine (2019b): Making Voices: The Gendering of Pitch Correction and the Auto-Tune Effect in Contemporary Pop Music. In: Journal of Popular Music Studies. 31/2. Pp. 63-84.

Rabiner, Lawrence (1977): On the Use of Autocorrelation Analysis for Pitch Detection. In: IEEE Transactions on Acoustics, Speech, and Signal Processing. 25/1. Pp. 24-33. DOI: 10.1109/tassp.1977.1162905.

Rodet, Xavier; Potard, Yves; Barriere, Jean-Baptiste (1984): The CHANT Project: From the Synthesis of the Singing Voice to Synthesis in General. In: Computer Music Journal. 8/3. Pp. 15-31. DOI: 10.2307/3679810.

Roseboro, Bronson (2019): The Vocaloid Phenomenon: A Glimpse into the Future of Songwriting, Community-Created Content, Art, and Humanity. Honor Scholar Thesis. DePauw University, Greencastle, IN, USA.

Sabo, Adriana (2019): Hatsune Miku: Whose Voice, Whose Body? In: INSAM Journal of Contemporary Music, Art and Technology. 1/2. Pp. 65-80.

Saewyc, Elizabeth M. (2017): Respecting Variations in Embodiment As Well As Gender: Beyond the Presumed 'Binary' of Sex. In: Nursing Inquiry. 24/1. DOI: 10.1111/nin.12184.

Schilt, Kristen; Westbrook, Laurel (2009): Doing Gender, Doing Heteronormativity. In: Gender & Society. 23/4. Pp. 440-464. DOI: 10.1177/0891243209340034.

Schröder, Marc (2009): Expressive Speech Synthesis: Past, Present, and Possible Futures. In: Affective Information Processing. Jianhua Tao; Tieniu Tan (eds.). London, UK: Springer London. Pp. 111-126. DOI: 10.1007/978-1-84800-306-4_7.

Skewes, Lea; Fine, Cordelia; Haslam, Nick (2018): Beyond Mars and Venus: The Role of Gender Essentialism in Support for Gender Inequality and Backlash. In: PLOS ONE. 13/7. Paper No. e0200921. DOI: 10.1371/journal.pone.0200921.

Štrkalj, Goran; Pather, Nalini (2021): Beyond the Sex Binary: Toward the Inclusive Anatomical Sciences Education. In: Anatomical Sciences Education. 14. Pp. 517-522. DOI: 10.1002/ase.2002.

Sutton, Selina J. (2020): Gender Ambiguous, not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. Proceedings of the 2nd Conference on Conversational User Interfaces (CUI '20). Bilbao, Spain. Pp. 1-8. Paper No. 11. DOI: 10.1145/3405755.3406123.

Tompkins, Dave (2011): How to Wreck a Nice Beach: The Vocoder from World War II to Hip-Hop: The Machine Speaks. Chicago, IL, USA: Melville House.

Vásárhelyi, Orsolya (2020): Computational and Relational Understanding of Gender Inequalities in Science and Technology. PhD thesis. Central European University, Budapest, Hungary.

West, Candace; Zimmerman, Don H. (1987): Doing Gender. In: Gender & Society. 1/2. Pp. 125-151. DOI: 10.1177/0891243287001002002.

West, Candace; Zimmerman, Don H. (2009): Accounting for Doing Gender. In: Gender & Society. 23/1. Pp. 112-122. DOI: 10.1177/0891243208326529.

West, Mark; Kraut, Rebecca; Chew, Han E. (2019): I'd Blush If I Could: Closing Gender Divides in Digital Skills Through Education. EQUALS and UNESCO. https://unesdoc.unesco.org/notice?id=p::usmarcdef_0000367416 [last accessed July 13, 2021].

Whittaker, Meredith; Alper, Meryl; Bennett, Cynthia L.; Hendren, Sara; Kaziunas, Liz; Mills, Mara; Ringel Morris, Meredith; Rankin, Joy; Rogers, Emily; Salas, Marcel; Myers West, Sarah (2019): Disability, Bias, and AI. AI Now Institute. https://ainowinstitute.org/disabilitybiasai-2019.pdf [last accessed July 13, 2021].

World Economic Forum (2018): Global Gender Gap Report 2018. http://www3.weforum.org/docs/WEF_GGGR_2018.pdf [last accessed July 13, 2021].

Yates, Kieran (2020): Why Do We Gender AI? Voice Tech Firms Move to Be More Inclusive. In: The Guardian. January 11. https://www.theguardian.com/technology/2020/jan/11/why-do-we-gender-ai-voice-tech-firms-move-to-be-more-inclusive [last accessed July 13, 2021].

Young, Samson (2012): A "Digital Opera" at the Boundaries of Transnationalism: Human and Synthesized Voices in Zuni Icosahedron's The Memory Palace of Matteo Ricci. In: Vocal Music and Contemporary Identities: Unlimited Voices in East Asia and the West. Christian Utz; Frederick Lau (eds.). New York, London: Routledge. Pp. 203-224. DOI: 10.4324/9780203078501.

Notes

[1] The assignment of sex is traditionally handled on the basis of externally expressed physical characteristics, namely genitals, at the time of birth (West & Zimmerman 2009).

[2] Research has also shown the binary view is not accurate for either sex (Štrkalj & Pather 2021; Saewyc 2017) or gender (Schilt & Westbrook 2009).

[3] More details about the technical basis of Auto-Tune can be found in Bridges (2020).

[4] Autotuning has become a generalized term for re-tuning, with or without noticeable timbral distortion. Since the early 2000s, it was not necessarily done exclusively with the Auto-Tune software as a wider range of products that offer pitch-correction technology began to be released, such as Celemony's Melodyne and Waves' Tune.

[5] The Oxford English Dictionary defines current usage of "camp" as "exaggerated, affected, over the top… Especially used with reference to the style or execution of a work of art or entertainment, or a dramatic performance", although previously it was associated "Especially of a man or his mannerisms, speech, etc.: flamboyant, arch, or theatrical, especially in a way stereotypically associated with an effeminate gay man." (OED 2021).

[6] The 2018 Global Gender Gap report by the World Economic Forum (2018) found that globally only 22% of AI professions were female-identifying, with the remaining 78% identified as male (the survey does not appear to have taken non-binary gender identity into account).

[7] The current Vocaloid product catalogue is available at https://www.vocaloid.com/en/products [last accessed July 26, 2021].

[8] Snapchat filters and audio effects can be explored at https://lensstudio.snapchat.com/guides/audio/audio-effect/ [last accessed July 26, 2021].