

Voice at NIME: a Taxonomy of New Interfaces for Vocal Musical Expression

Rébecca Kleinberger

MIT Media Lab, Massachusetts Institute of Technology,
Cambridge, MA,

Xiao Xiao

Léonard de Vinci Pôle Universitaire, Research Center,
92916 Paris La Défense, France,

Nikhil Singh

MIT Media Lab, Massachusetts Institute of Technology,
Cambridge, MA,

Akito van Troyer

Berklee College of Music, Electronic Production and
Design (EPD), Boston, MA

Abstract

We present a systematic review of voice-centered NIME publications from the past two decades. Musical expression has been a key driver of innovation in voice-based technologies, from traditional architectures that amplify singing to cutting-edge research in vocal synthesis. NIME conference has emerged as a prime venue for innovative vocal interfaces. However, there hasn't been a systematic analysis of all voice-related work or an effort to characterize their features. Analyzing trends in Vocal NIMEs can help the community better understand common interests, identify uncharted territories, and explore directions for future research. We identified a corpus of 98 papers about Vocal NIMEs from 2001 to 2021, which we analyzed in 3 ways. First, we automatically extracted latent themes and possible categories using natural language processing. Taking inspiration from concepts surfaced through this process, we then defined several core dimensions with associated descriptors of Vocal NIMEs and assigned each paper relevant descriptors under each dimension. Finally, we defined a classification system, which we then used to uniquely and more precisely situate each paper on a map, taking into account the overall goals of each work. Based on our analyses, we present trends and challenges, including questions of gender and diversity in our community, and reflect on opportunities for future work.

CCS Concepts

• **Applied computing** → Arts and humanities; • **General and reference** → Document types; • **Human-centered computing** → Human computer interaction (HCI).

Keywords

voice, vocal, singing, speech, analysis, synthesis, review, survey

1 Introduction

The history of voice technology is intimately connected to musical expression. From the atrium of Ancient Greek temples to the vaults of Byzantine and Gothic churches, the desire to enhance the singing voice has driven the study and engineering of acoustics in architecture [31, 84]. More recently, musical expression driving technological innovation can be seen in the history of audio recording. Both the first known audio recording by Martini in 1860 and the first demonstration of the phonograph by

Thomas Edison not only used human voices but human voices singing [47, 73] – respectively “Au clair de la lune” and “Mary had a little lamb”. Although tools for audio recording, playback and synthesis are now commonplace, the proliferation of such technologies led to profound shifts in our society as depicted by the controversial work of anthropologist Carpenter, who studies isolated populations being exposed to audio and video recording technology for the first time [93]. The intersection of voice technology with music can also have strong neurological and behavioral effects on individuals. Musically modulating the way someone hears their own voice, by manipulating pitch, formants, or using filters, has been associated with modifications in emotion [7, 27], semantic content [66], even disfluency regulation [67], highlighting the potential of musical voice technologies to create powerful transformative experiences.

Because of their musical affinities, artists and creative technologists have taken part in this impactful trend by creating vocal experiences that push the boundaries of technological innovation and musical expression. Notable examples include Alvin Lucier’s “I am sitting in a room” (1969) which uses the voice to deepen our perception of space. Far from being used merely as an input signal, the voice is here considered both in relation to the external architectural space, and in its connection with the self as Lucier had a severe stutter and used the inherent music of the room as a way to “smooth out any irregularities [his] speech might have” [75]. Another example is Laurie Anderson’s “Handphone Table” that challenges silence and materiality by transmitting a silent vocal signal by bone conduction directly to visitor’s ear [71]; and Machover’s “Philadelphia Voices” that creates a dialogue between diverse voices. In this piece, digital technologies allow the public to contribute their own voice to the final symphony [106].

The international conference on New Interfaces for Musical Expression began in 2001 as a workshop at the Conference on Human Factors in Computing Systems (CHI). After 20 years, the conference is well established and has reached a state of maturity. In recent years, the number of review articles presented at NIME has increased, indicating an introspective drive of the community to better understand its own roots, evolution, and possible future trajectories. Such review articles have targeted overarching values and questions in the research community such as evaluation [8], community [78], gender balance [119], sustainability [79] as well as types and presentation of work including sound installations [53], live performance [115], and score-centered NIMES [79].

Given the musical potential, high degrees of expressivity, and universality of the voice, the presence of voice-centered work at NIME is not surprising. Darwin even suggested that the voice is likely to be the first musical instrument [29]. But despite its familiarity, it can be challenging to clearly define what the voice



This work is licensed under a Creative Commons Attribution 4.0 International License.

NIME '22, June 28–July 1, 2022, Auckland, New Zealand

© 2022 Copyright held by the owner/author(s).

<https://doi.org/10.21428/92fbeb44.4308fb94>

is. Often considered solely in its acoustic manifestation, the voice is also studied in terms of its complex neurological basis, its mechanical and muscular processes, its biology in the context of clinical medicine, its cognitive and behavioral effects in psychology, etc. This is to say that the object “voice” can take a number of forms and is a fruitful ground for creative and innovative thinking around expressivity. Because of its diversity of backgrounds, skills and traditions, its freedom of not staying within the constraints of specific field boundaries, and its drive for expressivity, the NIME community appears uniquely equipped to explore the voice in creative and non-traditional ways. Thus, an encompassing categorization of Vocal NIMEs should start from a holistic and open-minded view of the voice.

Works about the voice have indeed been a consistent presence at NIME, and related academic communities, and previous attempts have been made at categorization. For example, d’Alessandro et al. present an overview encompassing decades of work on singing interfaces [35]. Focusing specifically on the NIME community, Reed and colleagues analyzed 20 papers on the control of the singing voice and proposed to distinguish between “voice as controller,” “controllers for vocal synthesis,” and “direct control” [97]. While this approach highlights the important distinction of the voice as an input versus output, we believe that the richness of Vocal NIMES can be described with more granularity and would benefit from including systems that invoke the voice not only in traditional singing contexts but within the full diversity of voice modalities.

Starting from a broad interpretation of what the voice can encompass, we present an attempt to gather, organize, classify, and categorize Vocal NIMES. We seek to describe diverse characteristics of existing Vocal NIMES and to inform future NIME works in which voice might play a central role. We aim to provide a qualitative yet structured assessment that is both critical and informative. To this end, we combined manual review and automatic natural language processing (NLP) in our effort to gather, extract, sort and categorize information to make sense of the field of Vocal NIMES.

We first describe our methodology for selecting papers and choices for types of demographic data to include, including methods of encoding gender and diversity, and present overall trends. We then present an initial investigation using unsupervised document clustering and topic modeling in an attempt to structure concepts for our corpus. Inspired by the emerging themes and informed by a thorough reading of all papers in our corpus, we propose a system of descriptors to encode the characteristics and key features of Vocal NIMES. Finally, integrating the authors’ intent and context of the work, we categorize articles according to their similarities. This analysis produced five categories, which we propose to use as a taxonomy to classify past and future Vocal NIMES. We conclude with insights, trends, and observations derived from our analysis.

2 Corpus and demographics

2.1 Corpus

In line with previous reflective studies of the NIME conference, we collected our corpus from the scientific proceedings of NIME from 2001 to 2021, with a total of 1,972 documents that include papers of all lengths. To identify publications with voice as a core component, we searched for all papers containing the words “voice,” “vocal,” “sing,” “singing” or “singer” at least once, yielding 521 documents. We then manually verified that the voice was a

central element of the work and discarded instances where voice is only used to refer to hardware (“voice-coil”, “singing bowl”) or is used metaphorically in reference to music tracks, melody lines, MIDI messages or voice-leading. We also excluded papers for which the voice was clearly a secondary part of the system, an afterthought, or used as an example on the same level as other instruments (“media such as voice, ambient sound and music”). We included six articles that described a collection of various works, a common format for early NIME publications, if at least one work was about the voice. The final corpus contained a total of 98 papers.

The temporal distribution of papers shows strong consistency as a minority but persistently present theme in the conference. Figure 1 plots for each year the number of Vocal NIME papers against the total number of publications that year. Ranging from 1 to 10 per year, Vocal NIME papers never represent more than 10% of the total proceedings but remained between 3 and 10% for the past decade.

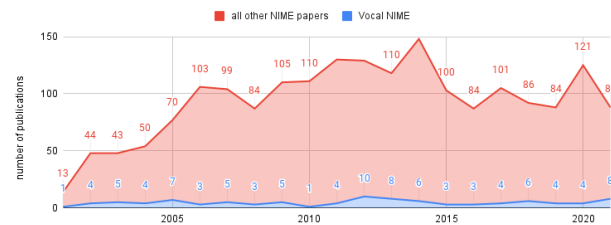


Figure 1: Figure 1: Number of Vocal NIME papers (in blue) compared to the total number of NIME publications (in red) from 2001 to 2021

2.2 Gender representation

We are interested in the gender representation of voice-related projects from our corpus because the voice is inherently connected with identity and gender. In terms of biology, the human voice is highly sexually dimorphic and can be seen as a strong marker of our hormonal identity. Social norms have also shaped the voice differently for men, women, and non-binary people throughout history [23, 89, 114].

We identified publications involving at least one woman author through internet searches on authors’ profiles, publicly available biographical information, and other texts from the author. This approach presents limitations as it does not account for non-binary authors or authors whose gender identity may not be reflected in publicly available information. To compare with the gender breakdown of NIME at large, we established the percentage of NIME publications including at least one woman author. The percentage from 2001 to 2018 was extracted from Xambó’s paper on gender balance [119], whose methodology we followed to derive the data for subsequent years. The results are shown in Figure 2. A Wilcoxon signed-rank test revealed that the observed difference between the ratio of papers with at least one woman author in our Vocal NIME corpus for each year ($Mdn.=38.78\%$) compared to the entire NIME corpus ($Mdn.=24.89\%$) are statistically significant, $p=.0089<.05$.

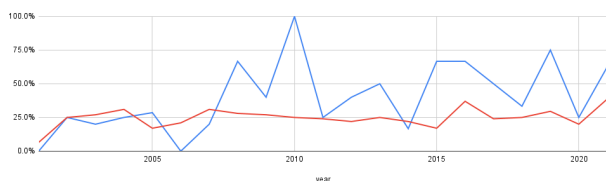


Figure 2: Figure 2: Percentage of papers with at least one woman author in the Vocal NIME corpus (in blue) and in the general NIME corpus (in red)

2.3 Geographical origins, accessibility, and diversity

Vocal traditions are a rich part of cultures around the world, and the voice presents an opportunity to introduce elements of geographic and cultural diversity at NIME. Our analysis first examined the geographic origins of Vocal NIMES compared to NIME at large based on the authors' country of affiliation. One limitation of this approach is that authors' original country of origin are not taken into account if they work in a different country. Figure 3 shows the distribution of non-unique authors by their affiliation country for Vocal NIMES on the left and for the entire NIME corpus on the right adapted from [44]. While at least one work from our corpus has come from five continents, the authors' affiliation countries are slightly less extensive than for NIME at large.

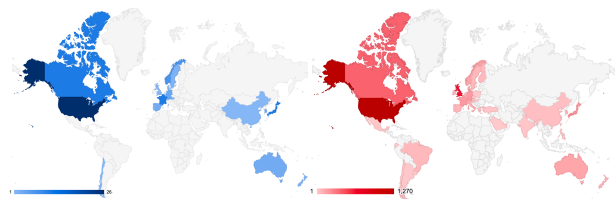


Figure 3: Figure 3: Distribution of non-unique authors by affiliation's country for Vocal NIMES (in blue) and for the entire NIME corpus (in red)

We also identified work that extends the understanding of the voice beyond traditional Western musical contexts and work that extends to underrepresented or marginalized populations. For each paper in our corpus, we determined whether some elements of non-western, indigenous, street, or folk cultures played a central role in some stage of the project, such as inspiration, or choice of practitioners. Looking for the inclusion of perspectives from underrepresented and otherly abled populations, we identified 17 works that make some reference to such themes. The first of these Vocal NIMES appeared only in 2007, and subsequent works maintain a marginal presence, averaging around 1 paper per year.

3 Unsupervised theme identification

To begin analyzing our corpus, we applied techniques from natural language processing to categorize papers and to suggest latent themes. We used two unsupervised machine learning methods to yield complementary insights: clustering to provide a top-down organization of the corpus, and topic models to identify latent topics. Results from both analyses were then interpreted to inform downstream manual annotation.

3.1 Preprocessing

From each paper's PDF file, we extracted text and addressed mis-read words (e.g. at line breaks). We then tokenized and lemmatized the corpus, removing words from a stopwords list including standard English stopwords, and voice-specific and NIME-specific terms that we identified as not useful to our analysis.

3.2 Document Clustering

Clustering attempts to automatically group data into discrete sets. Here, we expect these sets to correspond to a useful categorization of corpus papers. As a first step, we transformed our documents into a numerical feature representation: the popular term frequency-inverse document frequency (TF-IDF) technique. TF-IDF reflects how important a given word is to a document, and a document's TF-IDF feature vector has one such dimensional value for each word in the corpus vocabulary. In our case, this resulted in >800-dimensional feature vectors for our corpus. We applied k-Means clustering to our dataset, with k=8 clusters based on iterative manual inspection of documents and cluster coherence.

3.3 Topic Modeling

To explore possible latent themes in our corpus (in contrast to an imposed category structure), we used topic modeling via latent Dirichlet allocation (LDA). LDA is a generative model that uses latent variables to probabilistically explain similarity between data observations, in this case documents. Similar to the document clustering, we set the number of topics to 8 after an initial manual exploration.

3.4 Example

This automated analysis aimed to provide an overall view and initial insights to inform our survey approach. Rather than comprehensively survey these results, we detail results where relevant to our later analysis. For example, the largest cluster (N=28) in our cluster analysis contained several papers related to live performance. Inspecting word contributions by feature importance, we find the words *performer*, *sensor*, *gesture*, and *audience* to be most central to this cluster.

We later identify a category we call "Voice in performance/experience," which shares similar concepts. Conversely, smaller clusters contained central words like *mouth*, *tongue*, *microphone*, *shape*; which points to "Voice beyond audio", another category we later identify and define. In the sections that detail our manual characterization of papers in our corpus, we offer additional background information about aspects of the automated analysis.

4 Characteristics and descriptors

In this section, we use a hybrid deductive-inductive approach inspired by thematic analysis in identifying useful characteristics to describe Vocal NIME works. We deductively apply concepts from both prior work and our NLP-based automated analysis results, and inductively seek to discover details and other salient topics from the corpus. The pilot NLP-based analysis highlighted a series of important terms and possible ways to describe common themes and unique characteristics. However, these results could not be automatically converted to clear categories and characteristics (e.g. topic models are notoriously difficult to interpret), so we treated these results as surfacing useful concepts to consider in a methodical qualitative analysis procedure. One researcher

created an initial set of characteristics and categories by thoroughly examining and sorting papers in the corpus and reviewing NLP results. Two other researchers subsequently reviewed this categorization, and any disagreements were resolved through discussion. Following this, we propose five characteristics that can help describe Vocal NIMEs: Voice I/O, Behavioral Modality, Presentation Modality, Number of Voices, and Synchrony. Each of these five characteristics derives a series of possible attributes. In the following section, we define each characteristic and describe their attributes. Learnings from using this system to survey our corpus are presented in Figure 4.

4.1 Voice I/O

Prior work [97] has already highlighted the important distinction between systems where the voice is used as an input control vs systems generating and controlling a synthesized vocal output. Our topic model results surfaced some relevant terms, for example, topic 2 (top-4 words: synthesis, pitch, feature, user; suggesting analysis and control). Several related terms also had high TF-IDF scores (summed across documents), such as “controller” (#17 of all >800 words), “input” (#18), “microphone” (#33), and “output” (#51). By analyzing the corpus, it appears that this characterization goes further than a simple input vs output dichotomy. By considering NIME systems as black boxes, we noted six possible ways for the voice to interface with a system, which we define as descriptors:

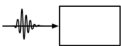
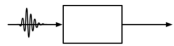
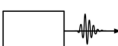
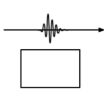
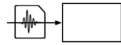
1. **voice analysis:** the voice is an interactive input to the system that is used solely to understand the signal 
2. **voice as control:** the voice is an interactive input to the system and its variation affect some non-voice-based output 
3. **voice synthesis:** some non-voice-based input affect or generate a synthesized voice signal 
4. **voice transformation/modulation/morphing:** a voice-related signal is both used as interactive input and output of the system 
5. **voice accompaniment:** voice is an important part of the experience but is neither an input nor output of the NIME, it exists alongside or in dialogue with the system 
6. **voice samples:** the vocal characteristics of the work consist in using pre-recorded vocal elements. 

Figure 4

Some of these descriptors can co-exist, for instance when the instruments operate in several different modes [100], when papers describe a collection of different projects [69], or in the case of live looping [9]

These descriptors allow us to define the context in which the voice is considered in the NIME community. In the case of voice as control, the voice often takes the shape of a given, natural

instrument is then extended. In the case of voice synthesis, it is seen as an intricate signal that we aim to teach machines to replicate. This I/O perspective also suggests different perspectives on expressivity depending on whether a NIME is used to shed light on or operate alongside a live performer (voice accompaniment), versus if the system inputs or outputs prerecorded static recordings (voice samples).

4.2 Behavioral modality

The focus of most NIME works is to enhance musical expression, in this sense, musical use of the voice through singing represents a large part of our corpus. However, our previous analysis also hints at the importance of the spoken voice, for example, topic 3 from our topic model (top-4 words: speech, synthesis, model, tract), with “speech” also being the 4th word in our corpus, by TF-IDF score (summed across all documents).

Looking at the corpus in more depth, we identified an even broader range of possible voice modalities denoting a rich, varied interpretation of vocal expressivity that includes vocalizations such as laughing and screaming. For simplicity, we segment the vocal modality used into three categories:

- (1) singing voice
- (2) spoken voice
- (3) other voice modalities

We noted a wide diversity of additional voice modalities from our corpus including breathing [3, 28, 58, 64, 65, 70, 82], tangible vibrations [4, 28, 57, 64, 65, 92], whispering [12, 17, 72], mouthing [77, 80], vocal spasm [37], beatboxing [45, 76], laughing [85], roaring [17], throat singing [118], screaming [72], humming [72], whistling [21], croaking [99], clicking [109], and buzzing [109].

4.3 Presentation modality

The NLP-based analysis highlights the embodiment of the voice. Topic 7 from our topic model strongly points to physical components of vocal production (top-4 words: mouth, tongue, microphone, shape). Even though most projects treat the voice as an audio signal, our corpus also contains work considering the voice as a complex, multimodal sensorial experience, such as laryngeal gestures, accompanying facial expressions, or muscle activation and movement. This denotes various levels of embodiment in the presentation of the vocal experience as an internal mechanical process. This internal feature of the voice is further shown through work considering the voice as purely internal by focussing on the inner voice. For the purpose of this work, we understand the term embodiment as the involvement of the body in the consideration of the voice. The body can be seen as secondary when the voice is only considered as a sound: the body can be in the foreground when the voice is considered as a movement, finally when the voice is considered as a neurological, silent phenomenon, it can be considered pre-embodied as the inner voice has been associated with neurological predictive signals for motor action preceding actual vocal action. From these considerations, we identified three descriptors for this characteristic:

The NLP-based analysis highlights the embodiment of the voice. Topic 7 from our topic model strongly points to physical components of vocal production (top-4 words: mouth, tongue, microphone, shape). Even though most projects treat the voice as an audio signal, our corpus also contains work considering the voice as a complex, multimodal sensorial experience, such as laryngeal gestures, accompanying facial expressions, or muscle

activation and movement. This denotes various levels of embodiment in the presentation of the vocal experience as an internal mechanical process. This internal feature of the voice is further shown through work considering the voice as purely internal by focussing on the inner voice. For the purpose of this work, we understand the term embodiment as the involvement of the body in the consideration of the voice. The body can be seen as secondary when the voice is only considered as a sound: the body can be in the foreground when the voice is considered as a movement, finally when the voice is considered as a neurological, silent phenomenon, it can be considered pre-embodied as the inner voice has been associated with neurological predictive signals for motor action preceding actual vocal action. From these considerations, we identified three descriptors for this characteristic:

- (1) the voice after leaving the body as a sound signal,
- (2) the voice in its embodied characteristics of movement, gesture or motor control
- (3) the voice in its pre-embodied form in the purely internal signal of an inner voice

These three possibilities often coexist, as in systems using internal voice-related signals (such as EEG, EMG, or camera vision of mouth shape) to control a synthesized vocal output.

When the voice is considered primarily as a sound signal, we noted a wide variety of acoustic features used by authors. Some of them are common, though at times appearing under different names or mathematical definitions such as Pitch/F0, Formants, Voicedness, or Vocal Timbre. Others are less common and not always consistent in their definition. This includes vibrato, loudness, amplitude, resonances, localization, intensity, vocal fold tenseness, MFCC Coefficients, LPC roots, timing, in-hale/exhale, decay, threshold, glottal waveform, inharmonicity, clarity, spectral centroid, spectral spread, spectral kurtosis, vocal effort, tongue shape, tenseness, and velum (nasality). This lack of consistency and common language highlights a potential need for unifying the terminology of voice parameters.

4.4 Number of voices

The voice is a fundamental instrument both for personal expression and for interpersonal communication. The voice as a tool for projection and expression is rich both on its own and in its potential to engage in dialogues. We marked a distinction between NIME works that consider one vocal signal and those that consider multiple voices. Again, results from our topic model pointed to distinct associations between solo and group activity: topic 1 (top-4 words: performer, sensor, gesture, audience) vs. topic 6 (top-4 words: participant, vibration, experience, group).

4.5 Synchrony

The level of synchrony also appears as a possible descriptor of our corpus. Indeed, although analysis or synthesis-based projects often aim at reducing processing delay for a seamless experience, other works use delay as an artistic tool (or even pre-recorded audio samples) to achieve new experiences. We distinguish three attributes for this characteristic:

- (1) Real-time processing (not introducing additional delay other internal processing lags)
- (2) Delayed (when the delay is intentional)
- (3) Use of voice samples (when the sound is decoupled in time from its generation)

Figure 4 summarizes the review of the Vocal NIME corpus according to the five characteristics.

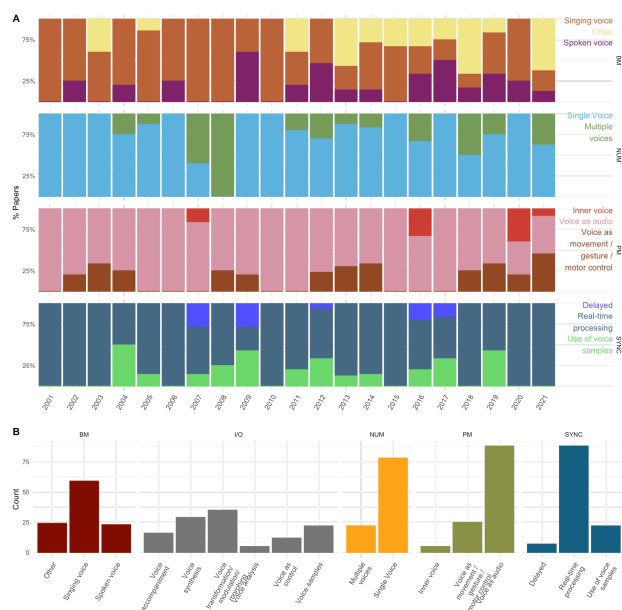


Figure 5: Figure 4: Distributions of paper characteristics (A) over time (normalized within each dimension), and (B) in aggregate. BM = Behavioral Modality, NUM = Number, PM = Presentation Modality, SYNC = Synchrony, I/O = Voice I/O. We exclude Voice I/O from (A) due to the number of characteristics.

5 Toward a taxonomy

Our characterization of Vocal NIMEs attempted to describe common features and different approaches. In this section, we go a step further by deriving a taxonomy that allows vocal NIMEs to be situated in a semantic map of the research space. Inspired by Ishii's pyramid [1], which postulates that research may be driven by pushing the boundaries of enabling technologies (bottom level), finding new applications (middle level), or in service of a vision (top level), our taxonomy also has three levels. These levels are similar in concept, but we find it more helpful to define each of them by the key question they ask about the system to categorize.

The top level which asks about the context, intention or vision behind the research, and a middle level that asks what the system does or enables people to do. The first two items of the top level (control of voice synthesis, voice as control) are taken directly from our previous categorization and are motivated by clear scientific goals. We established other new categories for similar but more artistic or exploratory work, as well as emerging applications.

- **Voice in performance:** Two different contexts emerged for papers with identical characterizations such as hand-based control of singing voice synthesis. Some are evaluated in terms of engineering feats and seek to push the boundaries of technical capabilities while others are designed for performance and evaluated as artistic experiences.

- **Voice training:** While only three papers belong to this category, training the voice is an age-old question with many nuances (e.g. breath control, rhythm, ear training), where new technologies can make contributions.
- **Voice Beyond Audio:** The framing of the voice as more than an audio signal is noteworthy in the context of expression and encompasses different angles including invoice voice, tangible vocal vibrations, and face/tongue movement tracking.
- **Non-Human Voices:** We observed a few works focusing on non-human voices, with most of them occurring in the past 5 years which may represent a rising trend. In two cases, the authors intentionally anthropomorphized to give “voice” to electrical components or socially interacting robots. In other cases, the papers consider the vocal experiences of other animals. A common theme of these explorations is the voice as a means to sensitize humans for other beings.

To disambiguate between works with similar motivation and design, we introduce a lower level that asks how the system is technically implemented. For instance, the large number of NIMEs that transfer the control of the voice to the hands can be differentiated based on the type of device controlled by the hands (e.g. augmented instrument, tablet, free-hand gestures). We present the result of this categorization situating the 98 papers of the corpus in Table 1.

Table 1 : Categorisation of Vocal NIMEs

To encourage easy exploration of this corpus through our taxonomy, we provide an interactive web-based tool [<https://nimevoice2022.vercel.app/>].

Embedded content.

6 Discussion and insights

As previously observed Vocal NIMEs have been a small but consistent part of the community over the years. Some works come from a small but dedicated contingent of researchers who consistently contribute new publications that explore similar themes. Others seem to be one-off explorations on some aspect of the voice with no follow-up in subsequent years. Here, we identify some opportunities for future research in terms of diversity-related themes and the implications of Vocal NIME research beyond NIME.

6.1 Diversity

Only seven papers in our corpus consider the voice beyond the context of traditional western music. In terms of extensions to ancient traditions, indigenous and marginalized communities, we noted references to Buddhist chanting, American street culture, ancient Egypt, Irish folk culture, Papua New Guinea, Tuvan throat singing, Chinese traditions, and Colombian migrants. This list represents geographic diversity, but there are even many more unique vocal traditions from around the world that can inspire future Vocal NIMEs.

In terms of work promoting inclusivity, we only noted two references to underrepresented populations in the deaf community and people with severe disabilities. Projects considering the voice beyond words and even beyond acoustic representations could be particularly relevant for people with disabilities, such as non-verbal children and adults, stroke survivors, or people who stutter.

Finally, one clear insight from our corpus demographics is the increased gender parity of Vocal NIMEs (38.7% with at least one

woman author) compared to NIME at large (24.9%). Although still far from gender balanced, the voice is one theme that can promote more collaborations between NIME researchers and researchers from fields with stronger women’s representation, such as Speech Language Pathology[16] and Linguistics [101].

Our analysis highlights the potential of the theme “voice” in inspiring researchers from various backgrounds. This may inform actions to increase equity and diversity in the community by bringing voice to the forefront or creating special interest groups around the theme of the voice.

6.2 Beyond NIME

As Vocal NIMEs present new analytic, interactive, and conceptual approaches to working with the voice, we explored how their impact might extend beyond the NIME community into other voice-related work. We gathered metadata from papers that cited those in our Vocal NIME corpus using Semantic Scholar (excepting 5 not-found papers; N=1169 from 920 unique papers). From each, we then extracted venue and field of study with the Semantic Scholar graph API, post-processing venues manually to aggregate (e.g. “NIME” and “NIME ‘07”).

References to Vocal NIMEs appear in a wide range of research venues, most commonly computer music and HCI venues including NIME (N=209), ICMC (N=38), the Computer Music Journal (N=31), SIGGRAPH ASIA (N=28), CHI (N=26), SIGGRAPH and Organized Sound (N=18 each), and TEI (N=17). ICASSP and INTERSPEECH, popular for speech and vocal processing, also appeared prominently (N=11 each). Other venues included multimedia (ACM Multimedia, N=9), movement (MOCO, N=6), acoustics (JASA, N=6), accessibility (ASSETS, N=2), psychology (Front. Psych., N=2), and particular works in further domains including robotics (IROS, HRI), affective computing (ACII, Trans. Aff. Comp.), and AI (AAAI, PRICAI).

We also looked at fields of study represented, via the Semantic Scholar API. The most frequently appearing field is Computer Science (N=854). Others include Engineering (N=133), Psychology (N=62), Art (N=30), Medicine (N=15), Sociology (N=13), Mathematics (N=13), Philosophy (N=7), Physics (N=2), Materials Science (N=1), and Geography (N=1). Although in practice we find these labels can often only be approximate(e.g. “Medicine” might reflect Physiology), we believe they broadly reflect the diversity of citing work.

Overall, vocal NIMEs have a wide-ranging impact on many near and distant academic fields. This suggests insights and methods from NIME can inform various aspects of the broader voice research landscape. We encourage authors interested in vocal NIMEs to consider the potential impact of their work in diverse disciplines.

7 Conclusion

This work contributes a taxonomy and descriptive system of key characteristics of Vocal NIMEs, based on a corpus we defined and analyzed with a mixed-methods approach. Our work offers a unified vocabulary to describe Vocal NIMEs, organizes trends in research themes, and identifies underexplored territories for future voice-related research. Our work has several limitations. For example, we only address research publications at NIME, not installations or performances. Our investigation suggests that the voice can be a bridge between diverse areas of research and artistic practice. We hope that our reflections can inspire new

Category (vision)	Subcategory (application)	Implementation Papers	
Control of voice synthesis / morphing	Hand-based control	tangible object	[2, 6, 20, 24–
		/ augmented	26, 112, 116,
		instrument	121, 123]
		screen/stylus	[5, 9, 32, 36, 38,
			50, 51, 85, 90]
		free	[43, 118, 120]
		movements (camera, BCI, EMG)	
		glove/wearable	[48, 54, 94, 113, 122]
		joystick	[62, 63, 108]
			[33, 61, 97, 107]
Voice as control	Other		
	Spoken		[10, 81, 81]
Voice in performance / experience	Sung		[19, 45, 59, 74]
	Hand-based control	tangible object / augmented instrument	[56, 87, 100, 117]
		free	[42]
		movements (camera, BCI, EMG)	
		glove/wearable	[14, 113]
			[60, 69, 91, 104]
		Voice Visualization	
		Social / Collaborative	[3, 4, 15, 18, 37, 39, 64, 72, 83, 95, 96]
		Other	[11–
			13, 28, 40, 41, 52, 55, 58, 82, 86, 98, 102, 103, 105]
			[21, 77, 80, 92, 111]
			[22, 68]
Voice beyond audio	Face/tongue tracking		[57, 65]
	Inner voice		
	Other / vibrations		
Voice training			[49, 70, 76]
Non-human voices			[17, 30, 34, 99, 109, 110]
Uncategorized			[46, 88]

Vocal NIMEs to enhance expression, connection, and well-being.

8 Ethics statement

None of the authors of this paper reported a conflict of interest. We addressed accessibility and inclusion by reviewing our corpus in light of gender balance and geographic diversity, as well as additionally discussing accessibility and references to non-western and marginalized cultures, which contributes to the open dialogue on inclusion within the NIME community. Our diversity- and inclusion-related findings are presented with the hope of encouraging more research in these directions. In terms of data privacy, only publicly available data were used in the analysis. This work didn't include human participants or animals.

9

References

- [1] [n.d.]. Ishii, H., Lakatos, D., Bonanni, L., & Labrune, J.-B. (2012). Radical atoms. *Interactions*, 19(1), 38–51. <https://doi.org/10.1145/2065327.2065337>.
- [2] Gabriela Bila Advincula, Don Derek Haddad, and Kent Larson. 2019. Grain Prism: Hieroglyphic Interface for Granular Sampling.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 284–285.
- [3] Ximena Alarcón Diaz, Victor Evaristo Gonzalez Sanchez, and Cagri Erdem. 2019. INTIMAL: Walking to Feel Place, Breathing to Feel Presence. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [4] Anders-Petter Andersson and Birgitta Cappelen. 2013. Designing Empowering Vocal and Tangible Interaction. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2013).
- [5] Maria Astrinaki, Nicolas d' Alessandro, and Thierry Dutoit. 2012. MAGE-A Platform for Tangible Speech Synthesis.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. Citeseer.
- [6] Maria Astrinaki, Nicolas D'Alessandro, Loic Reboursière, Alexis Moinet, and Thierry Dutoit. 2013. MAGE 2.0: New Features and Its Application in the Development of a Talking Guitar.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 547–550.
- [7] Jean-Julien Aucouturier, Petter Johansson, Lars Hall, Rodrigo Segnini, Lolita Mercadié, and Katsumi Watanabe. 2016. Covert Digital Manipulation of Vocal Emotion Alter Speakers' Emotional States in a Congruent Direction. *Proceedings of the National Academy of Sciences* 113, 4 (2016), 948–953.
- [8] Jeronimo Barbosa, Joseph Malloch, Marcelo M Wanderley, and Stéphane Huot. 2015. What Does "Evaluation" Mean for the NIME Community? *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2015).

- [9] Jeronimo Barbosa, Marcelo M Wanderley, and Stephane Huot. 2017. Exploring Playfulness in Nime Design: The Case of Live Looping Tools. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2017).
- [10] Alain Baumann and Rosa Sánchez. 2002. Interdisciplinary Applications of New Instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 1–5.
- [11] Ross Bencina, Danielle Wilde, and Somaya Langley. 2008. Gesture≈ Sound Experiments: Process and Mappings.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*. 197–202.
- [12] Andreas Bergsland. 2011. The Six Fantasies Machine—an Instrument Modelling Phrases from Paul Lansky’s Six Fantasies. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [13] Andreas Bergsland and Tone VAAse. 2012. Using a Seeing/Blindfolded Paradigm to Study Audience Experiences of Live-Electronic Performances with Voice.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*.
- [14] Mark A Bokowiec. 2011. V’OCT (Ritual): An Interactive Vocal Work for Body-coder System and 8 Channel Spatialization.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 40–43.
- [15] Graham Booth and Michael Gurevich. 2012. Collaborative Composition and Socially Constituted Instruments: Ensemble Laptop Performance Through the Lens of Ethnography.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [16] Steven Boyd and Nigel Hewlett. 2001. The Gender Imbalance Among Speech and Language Therapists and Students. *International journal of language & communication disorders* 36, S1 (2001), 167–172.
- [17] Courtney Brown, Sharif Razzaque, and Garth Paine. 2015. Rawr! A Study in Sonic Skulls: Embodied Natural History.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [18] Antonio Camurri, Corrado Canepa, Paolo Coletta, Barbara Mazzarino, and Gualtiero Volpe. 2008. Mapped Per Affetti Erranti: A Multimodal System for Social Active Listening and Expressive Performance.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 134–139.
- [19] Mark Cartwright and Bryan Pardo. 2014. Synthassist: An Audio Synthesizer Programmed with Vocal Imitation. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 741–742.
- [20] Alex Champagne, Bob Pritchard, Paul Dietz, and Sidney Fels. 2021. Investigation of a Novel Shape Sensor for Musical Expression. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. PubPub.
- [21] Deepak Chandran and Ge Wang. 2018. InterFACE: New Faces for Musical Expression.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [22] Nick Collins. 2007. Matching Parts: Inner Voice Led Control for Symbolic and Audio Accompaniment. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*. 220–223.
- [23] Sarah A Collins and Caroline Missing. 2003. Vocal and Visual Attractiveness Are Related in Women. *Animal behaviour* 65, 5 (2003), 997–1004.
- [24] Perry Cook. 2017. 2001: *Principles for Designing Computer Music Controllers*. Springer, 1–13.
- [25] Perry R Cook. 2005. Real-Time Performance Controllers for Synthesized Singing. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. Citeseer, 236–237.
- [26] Perry R Cook. 2009. Re-Designing Principles for Computer Music Controllers: A Case Study of SqueezeVox Maggie.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 218–221.
- [27] Jean Costa, Malte F Jung, Mary Czerwinski, François Guimbretière, Trinh Le, and Tanzeem Choudhury. 2018. Regulating Feelings During Interpersonal Conflicts by Changing Voice Self-Perception. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [28] Kelsey Cotton, Pedro Sanches, Vasiliki Tsaknaki, and Pavel Karpashevich. 2021. The Body Electric: A NIME Designed Through and with the Somatic Experience of Singing. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [29] Charles Darwin. 1964. *On the Origin of Species: A Facsimile of the First Edition*. Harvard University Press.
- [30] Gamhewage C De Silva, Tamara Smyth, and Michael J Lyons. 2020. A Novel Face-Tracking Mouth Controller and Its Application to Interacting with Bioacoustic Models. *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)* (2020).
- [31] Nico F Declercq and Cindy SA Dekeyser. 2007. Acoustic Diffraction Effects at the Hellenistic Amphitheater of Epidaurus: Seat Rows Responsible for the Marvelous Acoustics. *The Journal of the Acoustical Society of America* 121, 4 (2007), 2011–2022.
- [32] Samuel Delalez and Christophe d’ Alessandro. 2017. Vokinesis: Syllabic Control Points for Performative Singing Synthesis.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 198–203.
- [33] Liam Donovan and Andrew P McPherson. 2014. The Talking Guitar: Headstock Tracking and Mapping Strategies.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 351–354.
- [34] Paul Dunham, Mo H Zareei, Dale Carnegie, and Dugal McKinnon. 2021. Click: RAND# 2. An Indeterminate Sound Sculpture. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*. PubPub.
- [35] Christophe d’ Alessandro, Samuel Delalez, Boris Doval, Lionel Feugère, and Olivier Perrotin. 2019. Les Instruments Chanteurs. *Acoustique et Techniques: trimestriel d’information des professionnels de l’acoustique* 89 (2019), 36–43.
- [36] Nicolas d’ Alessandro and Thierry Dutoit. 2007. Handsketch Bi-Manual Controller: Investigation on Expressive Control Issues of an Augmented Tablet. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 78–81.
- [37] Nicolas D’Alessandro, Roberto Calderon, and Stefanie Müller. 2011. ROOM# 81-Agent-Based Instrument for Experiencing Architectural and Vocal Cues.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [38] Nicolas D’Alessandro, Christophe d’ Alessandro, Sylvain Le Beux, and Boris Doval. 2006. Real-Time CALM Synthesizer: New Approaches in Hands-Controlled Voice Synthesis.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 266–271.
- [39] Nicolas D’Alessandro, Aura Pon, Johny Wang, David Eagle, Ehud Sharlin, and Sidney S Fels. 2012. A Digital Mobile Choir: Joining Two Interfaces Towards Composing and Performing Collaborative Mobile Music.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [40] Gideon D’Arcangelo. 2002. Creating a Context for Musical Innovation: A NIME Curriculum. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*. 1–4.
- [41] Gideon D’Arcangelo. 2004. Recycling Music, Answering Back: Toward an Oral Tradition of Electronic Music.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*. Citeseer, 55–58.
- [42] Ayaka Endo, Takuma Moriyama, and Yasuo Kuhara. 2012. Tweet Harp: Laser Harp Generating Voice and Text of Real-time Tweets in Twitter.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [43] Cagri Erdem and Alexander Refsum Jensenius. 2020. RAW: Exploring Control Structures for Muscle-Based Interaction in Collective Improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. Birmingham City University, 477–482.
- [44] Stefano Fasciani and Jackson Goode. 2021. 20 NIMES: Twenty Years of New Interfaces for Musical Expression. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. PubPub.
- [45] Stefano Fasciani and Lonce Wyse. 2012. A Voice Interface for Sound Generators: Adaptive and Automatic Mapping of Gestures to Sound. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2012).
- [46] Stefano Fasciani and Lonce Wyse. 2013. A Self-Organizing Gesture Map for a Voice-Controlled Instrument Interface.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*. 507–512.
- [47] Patrick Feaster. 2010. Edouard-Leon Scott de Martinville: An Annotated Discography. *ARSC Journal*. 41, 1 (2010), 43–.
- [48] Sidney S Fels, Bob Pritchard, and Allison Lenters. 2009. ForTouch: A Wearable Digital Ventriloquized Actor. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 274–275.
- [49] Sam Ferguson. 2006. Learning Musical Instrument Skills Through Interactive Sonification. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*. 384–389.
- [50] Lionel Feugère and Christophe d’ Alessandro. 2014. Rule-Based Performative Synthesis of Sung Syllables. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2014).
- [51] Lionel Feugère and Christophe d’ Alessandro. 2013. Digitartic: Bi-Manual Gestural Control of Articulation in Performative Singing Synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 331–336.
- [52] Daniel Buner Formo. 2018. The Orchestra of Speech—a Speech-Based Instrument System. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*.
- [53] Valérian Fraisse, Catherine Guastavino, and Marcelo M Wanderley. 2021. A Visualization Tool to Explore Interactive Sound Installations. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2021).
- [54] Yoonchang Han, Jinsoo Na, and Kyogu Lee. 2012. FutureGrab: A Wearable Synthesizer Using Vowel Formants. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2012).
- [55] Anne Hege, Camille Noufi, Elena Georgieva, and Ge Wang. 2021. Instrument Design for The Furies: A LaptOpera. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*. PubPub.
- [56] Donna Hewitt and Ian Stevenson. 2003. E-Mic: Extended Mic-Stand Interface Controller. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. McGill University, Faculty of Music, 122–128.
- [57] Charles Holbrow, Elena Jessop, and Rebecca Kleinberger. 2014. Vocal Vibrations. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [58] Javier Jaimovich. 2016. Emovere: Designing Sound Interactions for Biosignals and Dancers.. In *Proceedings of the International Conference on New Interfaces*

- for Musical Expression, NIME.
- [59] Jordi Janer. 2005. Voice-Controlled Plucked Bass Guitar Through Two Synthesis Techniques. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 132–135.
 - [60] Ryo Kanda, Mitsuyo Hashida, and Haruhiro Katayose. 2009. Mims: Interactive Multimedia Live Performance System. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. Citeseer, 45–47.
 - [61] Hideki Kawahara, Hideki Banno, and Masanori Morise. 2004. Acappella Synthesis Demonstrations Using RWC Music Database: Application of Auditory Morphing Based on STRAIGHT. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 130–131.
 - [62] Loïc Kessous. 2002. Bi-Manual Mapping Experimentation, with Angular Fundamental Frequency Control and Sound Color Navigation. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. Citeseer, 1–2.
 - [63] Loïc Kessous and Daniel Arfib. 2003. Bimanuality in Alternate Musical Instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. Citeseer, 140–145.
 - [64] Chris Kiefer and Cécile Chevalier. 2018. Towards New Modes of Collective Musical Expression Through Audio Augmented Reality. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [65] Rebecca Kleinberger. 2018. Vocal Musical Expression with a Tactile Resonating Device and Its Psychophysiological Effects. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [66] Rebecca Kleinberger, Stefanakis George, and Sebastian Franjou. 2019. Speech Companions: Evaluating the Effects of Musically Modulated Auditory Feedback on the Voice. In *The 25th International Conference on Auditory Display, ICAD 2019*. ICAD 2019, UK, Northumbria University.
 - [67] Rebecca Kleinberger, Henrietta Marie Franca Kleinberger. 2020. *Vocal Connection Rethinking the Voice as a Medium for Personal, Interpersonal, and Interspecies Understanding*. Ph. D. Dissertation.
 - [68] Sang Won Lee, Georg Essl, and Mari Martinez. 2016. Live Writing: Writing as a Real-time Audiovisual Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 212–217.
 - [69] Golan Levin. 2005. A Personal Chronology of Audiovisual Systems Research. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [70] Yinmiao Li, Ziyue Piao, and Gus Xia. 2021. A Wearable Haptic Interface for Breath Guidance in Vocal Training. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2021).
 - [71] Yueh-Tuan Li and Wen-Shu Lai. 2013. Voice, Object and Listening in the Sound Installations of Laurie Anderson. *International Journal of Performance Arts and Digital Media* 9, 2 (2013), 347–361.
 - [72] Anders Lind. 2018. Animated Notation in Multiple Parts for Crowd of Non-Professional Performers. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [73] Greg Linder. 2001. *Thomas Edison*. Capstone.
 - [74] Alex Loscos and Thomas Aussenac. 2005. The Wahwactor: A Voice Controlled Wah-Wah Pedal. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. Citeseer, 172–175.
 - [75] Alvin Lucier. 1970. *I Am Sitting in a Room*. Vol. 1013. Lovely Music.
 - [76] Simon Lui. 2013. A Compact Spectrum-Assisted Human Beatboxing Reinforcement Learning Tool On Smartphone. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [77] Michael J Lyons, Michael Haehnel, and Nobuji Tetsutani. 2020. Designing, Playing, and Performing with a Vision-Based Mouth Interface. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2020).
 - [78] Adnan Marquez-Borbon and Paul Stapleton. 2015. Fourteen Years of NIME: The Value and Meaning of “Community” in Interactive Music Research. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [79] Raul Masu, Adam Pultz Melbye, John Sullivan, and Alexander Refsum Jensenius. 2021. NIME and the Environment: Toward a More Sustainable NIME Practice. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. PubPub.
 - [80] Alex McLean, Eunjo Shin, and Kia Ng. 2013. Paralinguistic Microphone. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [81] Alex McLean and Geraint A Wiggins. 2009. Words, Movement and Timbre. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 276–279.
 - [82] Yoichi Nagashima. 2003. Bio-Sensing Systems and Bio-Feedback Systems for Interactive Media Arts. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [83] Misako Nakamoto and Yasuo Kuhara. 2007. Circle Canon Chorus System Used to Enjoy a Musical Ensemble Singing “Frog Round”. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 409–410.
 - [84] AM Noxon. 2001. Understanding Church Acoustics. *Acoustic Sciences Corporation* (2001).
 - [85] Jieun Oh and Ge Wang. 2013. LOLOL: Laugh Out Loud On Laptop. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [86] Nicola Orio, Serge Lemouton, and Diemo Schwarz. 2003. Score Following: State of the Art and New Developments. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [87] Yongki Park, Hoon Heo, and Kyogu Lee. 2012. Voicon: An Interactive Gestural Microphone For Vocal Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [88] Ricardo Pedrosa and Karon E MacLean. 2009. Evaluation of 3D Haptic Target Rendering to Support Timing in Music Tasks. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. Citeseer, 19–24.
 - [89] Cecilia Pemberton, Paul McCormack, and Alison Russell. 1998. Have Women’s Voices Lowered Across Time? A Cross Sectional Study of Australian Women’s Voices. *Journal of Voice* 12, 2 (1998), 208–213.
 - [90] Olivier Perrotin and Christophe d’Alessandro. 2013. Adaptive Mapping for Improved Pitch Accuracy on Touch User Interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. Woon Seung Yeo, Kyogu Lee, Alexander Sigman, Haru Ji, Graham Wakefield, 186–189.
 - [91] Olivier Perrotin and Christophe d’Alessandro. 2014. Visualizing Gestures in the Control of a Digital Musical Instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 605–608.
 - [92] Cornelius Poepel, Jochen Feitsch, Marco Strobel, and Christian Geiger. 2014. Design and Evaluation of a Gesture Controlled Singing Voice Installation. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [93] Harald EL Prins and John Bishop. 2001. Edmund Carpenter: Explorations In Media & Anthropology. *Visual Anthropology Review* 17, 2 (2001), 110–140.
 - [94] Bob Pritchard and Sidney Fels. 2006. Grassp: Gesturally-realized Audio, Speech and Song Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. Citeseer, 272–276.
 - [95] Miguel A Ortiz Pérez, Benjamin Knapp, and Michael Alcorn. 2007. Diamair: Composing for Choir and Integral Music Controller. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 289–292.
 - [96] Chandrasekhar Ramakrishnan, Jason Freeman, and Kristjan Varnik. 2004. The Architecture of Auracle: A Real-Time, Distributed, Collaborative Instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 100–103.
 - [97] C Reed, Andrew McPherson, et al. 2020. Surface Electromyography for Direct Vocal Control. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [98] Paulo Maria Rodrigues, Luis Miguel Girão, and Rolf Gehlhaar. 2005. CyberSong. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 164–167.
 - [99] Margaret Schedel, Brian Smith, Robert Cosgrove, and Nick Hwang. 2021. RhumbLine: Plectrohyla Exquisita—Spatial Listening of Zoomorphic Musical Robots. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [100] Daniel Schlessinger. 2012. Concept Tahoe: Microphone Midi Control. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
 - [101] Angela Schrader et al. 2021. The State of Linguistics in Higher Education Annual Report 2020. *Linguistic Society of America* (2021).
 - [102] Alex Stahl and Patricia Clemens. 2010. Auditory Masquing: Wearable Sound Systems for Diegetic Character Voices. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 427–430.
 - [103] Dan Stowell, Mark D Plumbley, and Nick Bryan-Kinns. 2008. Discourse Analysis Evaluation Method for Expressive Musical Interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 81–86.
 - [104] Robyn Taylor, Daniel Torres, and Pierre Boulanger. 2005. Using Music to Interact with a Virtual Character. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 220–223.
 - [105] Enrique Tomás, Thomas Gorbach, Hilda Tellioglu, and Martin Kaltenbrunner. 2021. Embodied Gestures: Sculpting Energy-Motion Models into Musical Interfaces. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2021).
 - [106] Akito van Troyer. 2013. Constellation: A Tool for Creative Dialog Between Audience and Composer. In *10th International Symposium on Computer Music Multidisciplinary Research*.
 - [107] Satvik Venkatesh, Edward Braund, and Eduardo Reck Miranda. 2020. Designing Brain-computer Interfaces for Sonic Expression. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2020).
 - [108] Benjamin Vigoda and David Merrill. 2007. Jamioki-Purejoy: A Game Engine and Instrument for Electronically-Mediated Musical Improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 321–326.
 - [109] Nathan Villcaña-Shaw, Dale A Carnegie, Jim Murphy, and Mo Zareei. 2021. Speculator: Visual Soundscape Augmentation of Natural Environments. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. PubPub.

- [110] Nathan Villicaña-Shaw, Spencer Salazar, and Ajay Kapur. 2018. Mechatronic Performance in Computer Music Compositions.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*. 413–418.
- [111] Florian Vogt, Graeme McCaig, Mir Adnan Ali, and Sidney S Fels. 2002. Tongue'n'Groove: An Ultrasound Based Music Controller.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*. 60–64.
- [112] Johny Wang, Nicolas D'Alessandro, Sidney S Fels, and Bob Pritchard. 2011. SQUEEZY: Extending a Multi-touch Screen with Force Sensing Objects for Controlling Articulatory Synthesis.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 531–532.
- [113] Johny Wang, Nicolas D'Alessandro, Sidney S Fels, and Robert Pritchard. 2012. Investigation of Gesture Controlled Articulatory Vocal Synthesizer Using a Bio-Mechanical Mapping Layer.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [114] Stephanie Watson. 2019. The Unheard Female Voice: Women Are More Likely to Be Talked over and Unheeded. But SLPs Can Help Them Speak up and Be Heard. (2019).
- [115] Travis West, Baptiste Caramiaux, Stéphane Huot, and Marcelo Wanderley. 2021. Making Mappings: Design Criteria for Live Performance. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* (2021).
- [116] J Wu, Nathan Weitzner, Yoo Yeh, Jonathan Abel, Romain Michon, and Matthew Wright. 2015. Tibetan Singing Prayer Wheel: A Hybrid Musical-Spiritual Instrument Using Gestural Control. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [117] Jiayue Cecilia Wu, Madeline Huberth, Yoo Hsiu Yeh, and Matthew Wright. 2016. Evaluating the Audience's Perception of Real-time Gestural Control and Mapping Mechanisms in Electroacoustic Vocal Performance.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 206–211.
- [118] J Cecilia Wu, Mark Rau, Yun Zhang, Yijun Zhou, and Matthew James Wright. 2017. Towards Robust Tracking with an Unreliable Motion Sensor Using Machine Learning.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*.
- [119] Anna Xambo Sedo. 2018. Who Are the Women Authors in NIME?—improving Gender Balance in NIME Research. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. Virginia Tech Blacksburg, VA.
- [120] Xiao Xiao, Grégoire Locqueville, Christophe d' Alessandro, and Boris Doval. 2019. T-Voks: The Singing and Speaking Theremin. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 110–115.
- [121] Kazuhiko Yamamoto and Takeo Igarashi. 2016. Livo: Sing a Song with a Vowel Keyboard. *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME* 3 (2016), 460–468.
- [122] Tomoko Yonezawa, Noriko Suzuki, Kenji Mase, and Kiyoshi Kogure. 2005. Handysinger: Expressive Singing Voice Morphing Using Personified Hand-Puppet Interface. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 121–126.
- [123] Fou Yoshimura and Kazuhiro Jo. 2019. A" Voice" Instrument Based on Vocal Tract Models by Using Soft Material for a 3D Printer and an Electrolarynx.. In *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME*. 411–412.