

Forum on Transcription

JASON STANYEK – convener

FERNANDO BENADON, TARA BROWNER, PARAG CHORDIA, ANNE DANIELSEN, EMILIA GÓMEZ,
SUMANTH GOPINATH, DAI GRIFFITHS, KIRI MILLER, RACHEL MUNDY, JENNIFER ROTH-GORDON,
DAVID ROTHENBERG, MICHAEL TENZER – participants

Introduction

Fifty years ago, in September 1964, the journal of a young Society for Ethnomusicology published the ‘Symposium on Transcription and Analysis: A Hukwe Song with Musical Bow’, a text that has since become a cornerstone within the ethnomusicological corpus.¹ Drawn from a session organized by Nicholas M. England for SEM’s November 1963 annual conference, the Symposium was built from what England called the ‘devoted labors’ of Robert Garfias, Mieczyslaw Kolinski, George List, and Willard Rhodes, four key figures in the emerging discipline, each of whom contributed an idiosyncratic transcription of a performance by a San bow player named Kafulo that England had recorded in September 1959 in what is now northeastern Namibia.² Charles Seeger served as the ‘Chairman-Moderator’ and provided a report that included a ‘synoptic view’ of the four transcriptions.³ [See Figure 1.]

The Symposium stands as a monument to musical transcription – or what Seeger in his report calls, pertinently (though somewhat reductively), the ‘visual documentation of sound-recording’.⁴ It throws into relief perspectives that, fifty years on, are still relevant – almost axiomatically so – for scholars who produce and analyse transcriptions of musical and sonic events. Amongst these perspectives are the following: *total accuracy is impossible* (Garfias: ‘No system of transcription, mechanical or otherwise, can preserve all of a musical example accurately’);⁵ *the sonic is recalcitrant to inscription* (Rhodes: ‘I was keenly conscious of the minute variations of pitch, dynamics, and rhythm of both the bow and the voice, but I found them so small as to elude accurate notation with our present means’);⁶ *a transcription can be full or partial* (Kolinski: ‘[whether the recording] should be transcribed in extenso or whether it suffices to select a representative portion of it’);⁷ *transcription is contingent* (List: ‘In transcribing a musical fabric as complex as the one under consideration I should probably change my opinion concerning certain details on almost

<jason.stanyek@music.ox.ac.uk>

- 1 Nicholas M. England, et al., ‘Symposium on Transcription and Analysis: A Hukwe Song with Musical Bow’, *Ethnomusicology* 8/3 (1964), 223–277.
- 2 England, et al., ‘Symposium on Transcription and Analysis’, 223.
- 3 Charles Seeger, ‘Report of the Chairman-Moderator’, *Ethnomusicology* 8/3 (1964), 274–275.
- 4 Seeger, ‘Report of the Chairman-Moderator’, 277.
- 5 Robert Garfias, ‘Transcription I’, *Ethnomusicology* 8/3 (1964), 233.
- 6 Willard Rhodes, ‘Transcription IV’, *Ethnomusicology* 8/3 (1964), 271.
- 7 Mieczyslaw Kolinski, ‘Transcription II’, *Ethnomusicology* 8/3 (1964), 241.

SYNOPTIC VIEW OF THE FOUR NOTATIONS

Figure 1 ‘Hukwe Bow Song: Synoptic View of the Four Notations’. The staves are labelled with the initial of the last name of each transcriber (G = Robert Garfias; R = Willard Rhodes; L = George List; K = Mieczyslaw Kolinski). Originally published in Charles Seeger, ‘Report of the Chairman-Moderator’, *Ethnomusicology* 8/3 (1964), 274. Used with permission of the Society for Ethnomusicology.

every hearing’);⁸ *the ear cannot capture all* (Rhodes: ‘The human ear has its limitations’);⁹ *transcription is a threshold function of human perception* (Garfias: ‘A twentieth of a second is generally agreed to be the limit of human perception of change, difference or error, the so-called “Just Noticeable Difference” factor. Any event occurring with a smaller degree of variation would not be perceived by normal humans’);¹⁰ *audibility is helped by slowing a recording down* (List: ‘it is extremely difficult to transcribe the bow part, and to some extent the vocal part, at the original speed. . . I therefore reduced the speed by changing the electric supply to 53 cycles rather than 60’);¹¹ *machines augment human perception* (Garfias: ‘mechanical aids can be a valuable extension of one’s native perception’);¹² *machines hear differently*

8 George List, ‘Transcription III’, *Ethnomusicology* 8/3 (1964), 253.

9 Rhodes, ‘Transcription IV’, 272.

10 Garfias, ‘Transcription I’, 240.

11 List, ‘Transcription III’, 252.

12 Garfias, ‘Transcription I’, 240.

than humans (Garfias: ‘the body of each tone has been indicated as a straight line, although, in fact, mechanical transcription with equipment such as the Seeger Melograph would undoubtedly show a melodic line of constantly fluctuating pitch. It seemed more important to indicate the different types of entrances and connections between pitches than the minor fluctuations of pitch which are, in any event, not really discriminated by the ear’);¹³ *transcription is a performance marked by repetition* (Rhodes: ‘The transcription is the result of many repeated “listenings” over a period of time’);¹⁴ *transcription involves measurement* (Garfias: ‘a Mark II Brush Recorder . . . transcribed these beats on graph paper traveling at a speed of 50 millimeters per second. The distance between the bow strokes was measured with the aid of an abstract ruler with 50 sub-divisions to the inch’);¹⁵ *notational systems are modifiable* (List: ‘It will be noted that in the present transcription I have used various signs and symbols as needed to indicate pitches differing from those of equal temperament, indefinite pitches, etc.’);¹⁶ *context matters* (Garfias: ‘when a single example is studied outside its cultural and musical context, it becomes very difficult to isolate the pertinent elements’).¹⁷

The Symposium’s fifty-five pages are filled with an abundance of vital observations and perspectives (and there are many more than those listed in the preceding paragraph). Clearly, a number of these were anticipated in earlier theoretical publications on transcription.¹⁸ Already by the first decade of the twentieth century, for example, Abraham and Hornbostel listed the following as one ‘technique of transcription’: ‘The speed of the playback should be reduced in working with unique, rapid, complex or highly melismatic passages’.¹⁹ But the Symposium also foreshadowed what was to come; many of its perspectives get taken up again and again in the rather replete literature on transcription that emerged in its wake.²⁰ In this regard, the brief pause I’ve taken here on the Symposium is worthwhile,

13 Garfias, ‘Transcription I’, 233.

14 Rhodes, ‘Transcription IV’, 265.

15 Garfias, ‘Transcription I’, 239.

16 List, ‘Transcription III’, 258.

17 Garfias, ‘Transcription I’, 238.

18 See, for example, Otto Abraham and Erich M. von Hornbostel, ‘Suggested Methods for the Transcription of Exotic Music’, *Ethnomusicology* 38/3 (1994 [1909–10]), 425–456; A. M. Jones, ‘On Transcribing African Music’, *African Music* 2/1 (1958), 11–14; Charles Seeger, ‘Prescriptive and Descriptive Music Writing’, *Musical Quarterly* 44/2 (1958), 184–195.

19 Abraham and Hornbostel, ‘Suggested Methods’, 445.

20 The post-1964 literature on transcription in ethnomusicology and popular music studies is immense but, more than 20 years after its publication, Ter Ellingson’s comprehensive book chapter ‘Transcription’ (in *Ethnomusicology: An Introduction*, ed. Helen Myers (New York: W.W. Norton & Company, 1993), 110–52) remains the single most useful English-language overview of the history and theory of ethnomusicological transcription. Though less historically oriented than Ellingson’s piece, Peter Winkler’s incisive ‘Writing Ghost Notes: The Poetics and Politics of Transcription’ (in *Keeping Score: Music, Disciplinarity, Culture*, ed. David Schwartz, Anahid Kassabian, and Lawrence Siegel (Charlottesville: University of Virginia Press, 1997), 169–203) serves a similar role with regard to transcription’s function within popular music studies. Other key English-language sources are: Kofi Agawu, *Representing African Music: Postcolonial Notes, Queries, Positions* (New York: Routledge, 2003); Marcia Herndon, ‘Analysis: The Herding of Sacred Cows’, *Ethnomusicology* 18/2 (1974), 219–62; Nazir Jairazbhoy, ‘The “Objective” and Subjective View in Music Transcription’, *Ethnomusicology* 21/2 (1977), 263–73; Marin Marian-Balasa, ‘Who Actually Needs Transcription?’

and not just as a commemoration of its half-century anniversary; the text that England and his colleagues generated remains a crucial index of what remains a core concern of a particular brand of music studies, a concern that goes beyond transcription's varied acts and products: *the representation of the musical or sonic event*.

Fifty years after the Symposium, how do the practices, products, and politics of transcription fit into the ever-changing landscape of ethnomusicology? Can we still call transcription, as England did in 1964, one of ethnomusicology's 'most important tools of the trade'?²¹ And, looking beyond the confines of the ethnomusicological to the broader and increasingly interdisciplinary realm of music studies, can we make enquiries into transcription's diverse use values, its varied braids with positivism, its evidentiary and credentializing dimensions, its sensual politics, its ethics, its techniques, its audiences, its pedagogies, its patrimonies? What is transcription's status for scholars of popular music? For music theorists? For musicologists? For those working in the interstices between music studies and animal studies, or music studies and computer science? The 1964 Symposium sparked a substantive, half-century-long conversation on transcription – in which directions has that conversation gone?

The Fates, 'Devoted Labors', and Allegories of Transcription

Fifty years is a long time in academic life, certainly long enough for once-vibrant 'tools of the trade' to ossify, or be rendered irrelevant or passé. Indeed, over the past decade or so, a number of observers have noted a declining role for transcription within some disciplines of music studies. In his polemical article from 2005, 'Who Needs Transcription? Notes on the Modern Rise of a Method and the Postmodern Fall of an Ideology', Marin Marian-Balasa answers his own stark question – 'who still needs, nowadays, to make and refer to musical transcriptions?' – with a stark answer: 'Indeed, many very-ethnomusicological books lack any transcribed musical text, and there are books with just a few scanned, music-related photos or even scores just inserted for their decorative quality rather than their musicological relevance'.²² Bruno Nettl makes a similar point in the 2005 edition of his book *The Study of Ethnomusicology*: 'By the twenty-first century, ethnomusicologists were doing far less transcribing than before; they could, after all, give their audience recordings, and one no longer had to prove by transcribing that the strange sounds were proper music'.²³

Notes on the Modern Rise of a Method and the Postmodern Fall of an Ideology', *The World of Music* 47/2 (2005), 5–29; Bruno Nettl, 'I Can't Say a Thing until I've Seen the Score: Transcription', in *The Study of Ethnomusicology: Thirty-One Issues and Concepts* (Urbana: University of Illinois Press, 2005), 74–91; Kay Kaufman Shelemay, 'Notation and Oral Tradition in Africa', in *The Garland Handbook of African Music*, ed. Ruth Stone (New York: Garland Publishing, 1999), 24–43. A special issue on transcription of *The World of Music* (47/2; 2005) contains a number of rich essays. But these represent just the tip of the iceberg. There is also a vibrant and rapidly expanding literature on transcription in the interdisciplinary field of music information retrieval (MIR); I indicate some key MIR sources in footnote 66.

21 England, 'Symposium on Transcription and Analysis', 223.

22 Marian-Balasa, 'Who Actually Needs Transcription?', 5.

23 Nettl, 'I Can't Say a Thing until I've Seen the Score', 85. It should be pointed out that the 1964 issue of the journal *Ethnomusicology* containing the 'Symposium' was accompanied by a supplemental recording of Kafulo's performance of the bow song that England had recorded in 1959.

My own knee-jerk sense is that Marian-Balasa and Nettl are correct (as are the many others who have proffered similar observations):²⁴ transcription is not as central as it once was, either as a technique for producing visual representations of musical or sonic events, or as a fundamental methodology undergirding the aspirations of at least some of the various disciplines that make up music studies. As I relate anecdotally in my conversation with Tara Browner and Michael Tenzer below, transcription's demise has seemed rather pronounced among the graduate students I've worked with, the vast majority of whom have, for the past eleven years, consistently avoided giving even the most cursory glance to the transcriptions that appear in our required seminar readings. Or, another anecdote, this one involving one of the few graduate students at the University of Oxford who actually still *does* use transcription heavily in his research: when the editors of a prominent journal recently rejected a transcription-laden paper written by this student, they did so by saying, with intense euphemism, that the journal would be 'unable to use the piece since the publication is moving away from *musicological readings* at this time' (emphasis mine). Perhaps Timothy Rice is right when he highlights the 'shift around 1980 from scientific theory and method toward hermeneutics and social theory, a shift marked by, indeed causing, a concomitant decline in publications devoted solely to the methods of music analysis, transcription, and fieldwork'.²⁵

On the other hand, even a quick perusal of recent publications in ethnomusicology and popular music studies (to cite two fields with particular proclivities for transcription) reveals that many scholars are still compelled to include transcriptions as part of their written research. Gabriel Solis noticed this when he wrote in 2012 that 'transcription was, *and is*, a common step in coming to a fixed text for analysis'.²⁶ Indeed, I don't think it's a stretch to say that early-twenty-first-century academic music books and journals continue to reserve an important spot for transcription within their visual economies. But to what extent? One possible answer to this question can be found in Tables 1 and 2 which provide a graphic representation of the incidence of transcription in articles appearing in two leading music journals, *Ethnomusicology* and *Popular Music*. The graphs are undeniably crude: they don't, for example, take account of subject matter (is transcription more prevalent in studies of certain places, styles, or genres?); of generational cohort (are younger scholars less inclined to go the transcription route than more advanced scholars?); of region (do scholars in particular parts of the world show a greater allegiance to transcription than those in others?); of editors and peer-reviewers (do particular editors or peer-reviewers have a greater propensity for soliciting and accepting articles containing transcription?); or of a host of other factors. Nevertheless, the data are instructive. It is fairly apparent, for instance, that, over a

24 For another example see Jeff Titon's 'Knowing Fieldwork', in *Shadows in the Field: New Perspectives for Fieldwork in Ethnomusicology*, ed. Gregory Barz and Timothy J. Cooley (New York: Oxford University Press, 2008), 25: 'Not long ago, musical transcription was the distinguishing mark of our discipline. . . . [t]oday it is not transcription but fieldwork that constitutes ethnomusicology'.

25 Timothy Rice, 'Ethnomusicological Theory', *Yearbook for Traditional Music* 42 (2010), 104–105.

26 Gabriel Solis, 'Thoughts on an Interdiscipline: Music Theory, Analysis, and Social Theory in Ethnomusicology', *Ethnomusicology* 56/3 (2012), 543 (emphasis mine).

Table 1 Articles with transcription in the journal *Ethnomusicology* from 1964–2013 (volumes 8–57). The upper graph shows the number of articles containing transcription relative to the total number of articles in each volume. The lower graph shows the percentage of articles containing transcription per volume. Data compiled by Johannes Snape. Graphs created by Maria Luiza Gatto using the statistical software package Stata.

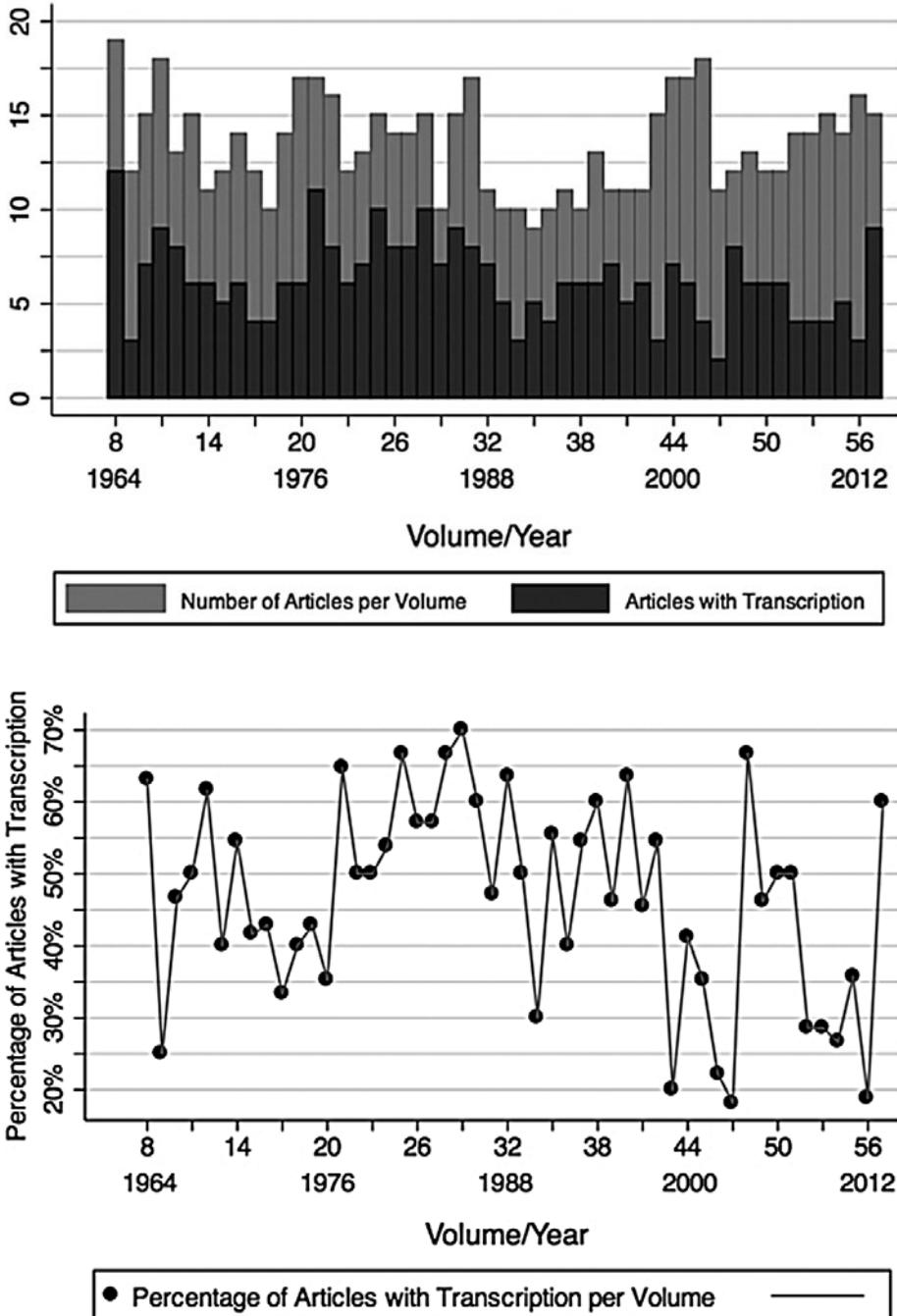
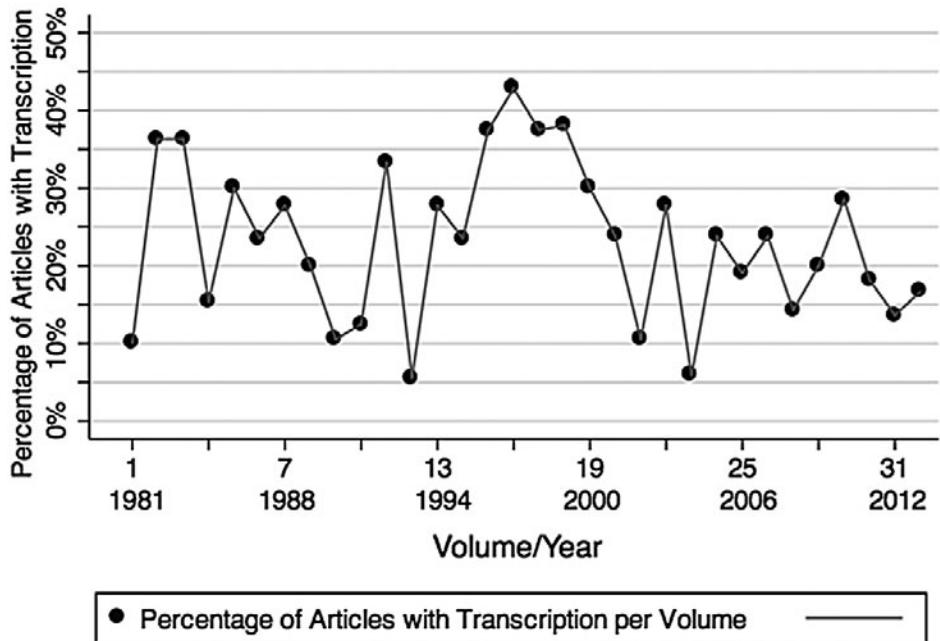
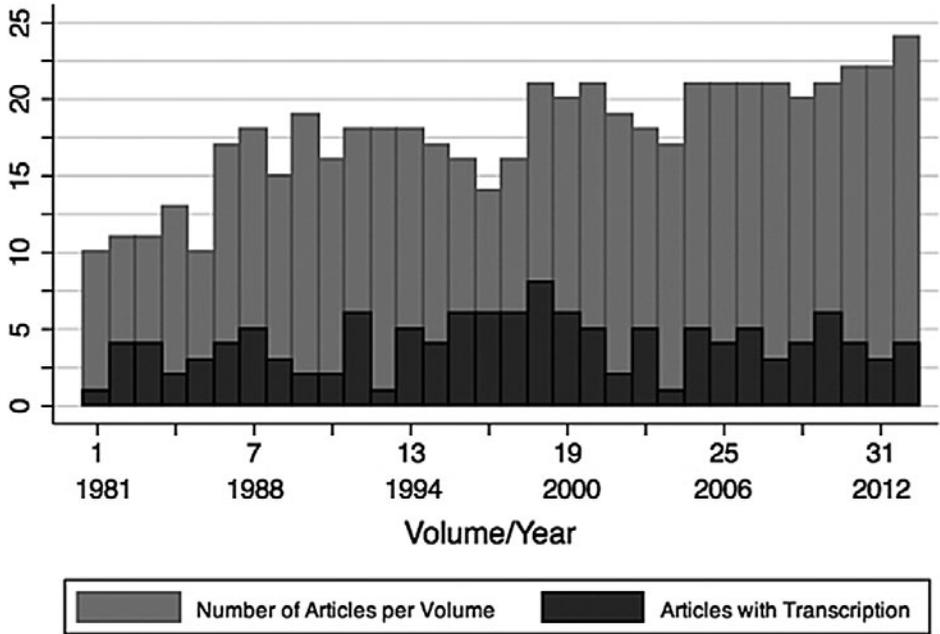


Table 2 Articles with transcription in the journal *Popular Music* from 1981–2013 (volumes 1–32). The graph on the left shows the number of articles containing transcription relative to the total number of articles in each volume. The graph on the right shows the percentage of articles containing transcriptions per volume. Data compiled by Johannes Snape. Graphs created by Maria Luiza Gatto using the statistical software package Stata.



fifty-year period in *Ethnomusicology* and a thirty-two-year period in *Popular Music*, each journal's commitment to publishing texts containing transcriptions seems to have remained intact. In terms of sheer quantity, there are ebbs and flows and more sudden upturns or downturns – and one can notice, perhaps, a distinct overall decrease in the number of articles containing transcriptions in *Ethnomusicology* – but, by and large, transcribed examples continue to have an important presence in the research published in these two key journals.

Clearly, transcription is not simply a form of visual representation and, while quantifying its presence within music-oriented publications is suggestive, this exercise ultimately tells us more about the profusion of transcription's products than it does about the anatomies of its practices. Transcription is also an *act*, a very specific mode of performance. It is a form of 'devoted labor', as Nicholas England named it in 1964,²⁷ calling attention to the fact that transcription is often conceptualized as hard work that involves an almost ritualistic form of repetition, a heightened devotional relation to music and sound. In *Representing African Music*, Kofi Agawu, for example, links transcription explicitly to work, suggesting that '[t]he archive of African music would be greatly impoverished without the labors of these workers'.²⁸ Transcription is not considered just any kind of work though: it is a kind of labour that is often framed as *laborious*. In his recent *Ethnomusicology: A Very Short Introduction* (2014), Timothy Rice notes how from the 1950s to the mid-1970s 'the effort to understand the structural elements of music was poured into laborious, handwritten descriptive transcriptions of the sound recordings ethnomusicologists had made in the field'.²⁹ Mantle Hood makes a similar point in the chapter on transcription in his book *The Ethnomusicologist* (one of the discipline's fundamental texts). We hear of transcription as a 'tedious chore' and also of the 'countless repetitions required in making a transcription' (which Hood sometimes frames as 'music dictation', a practice that necessitates the 'exercising ... of muscles').³⁰ For Peter Winkler, his attempt to transcribe Aretha Franklin's 'I Never Loved a Man (The Way I Love You)' involved 'starts and stops and obsessive repetitions of tiny scraps of sound' which served to bring him 'in touch with the music'.³¹ This is the devoted labour of transcription as communion with sound.

The heavy focus on the labours of the transcriber has often been at the expense of a sensitive accounting of the forms of labour on either side of the transcriptional act: the creative labour of musicians and the hermeneutic labour of the users of transcriptions. For instance, the scopic imperatives of transcription can sometimes serve to downplay the contributions of performers, whose fecund corporeal acts often disappear under the myopic and steely empirical gaze of the transcriber. A case in point is the virtual non-presence of the bow player Kafulo in the four contributions that are at the centre of the 1964 Symposium. England does describe Kafulo's performance in his introduction (which, pertinently, *does not* include any transcription) but, in his 'instructions' to the four transcribers, England

27 England, 'Symposium on Transcription and Analysis', 223.

28 Agawu, *Representing African Music*, 51.

29 Timothy Rice, *Ethnomusicology: A Very Short Introduction* (Oxford: Oxford University Press, 2014), 40.

30 Mantle Hood, 'Transcription and Notation', in *The Ethnomusicologist* (New York: McGraw-Hill, 1971), 59; 54.

31 Winkler, 'Writing Ghost Notes', 188.

says something revealing: ‘You will each bring up the points you think important for an understanding of what Kafulo (that is the old man’s name) is doing with regard to the acoustical phenomena themselves, and their organization into a musical organism’. In this instance – and indeed with regard to almost all transcription of the sort we’re discussing here – what takes precedence is the ‘musical organism’, itself the congealed labour of the musician, and now the object of the inscriptional labour of the transcriber. Also often given short shrift are the very particular modes of labour necessitated by the use of a transcription. This is revealed – to cite one example – in the frequent omission of tempo markings from the musical transcriptions that appear in academic publications. The very nature of a user’s temporal relationship with the transcribed music is obscured – evidence, perhaps, that the reconstitution of musical performance is not the primary use value of transcription.

What *is* transcription’s primary use value then? In many senses, the ‘devotional labor’ described above is inextricably linked to the disciplinary politics of the academy; as Peter Winkler puts it, transcription is very much about ‘the production of items of exchange in the academic economy’.³² Bruno Nettl looks at transcription from a related angle, pointing out that ‘a good many ethnomusicologists are emotionally tied to the sound of music and get much pleasure in their work from transcribing; and having made a transcription gives them a certain sense of *direct ownership and control* over the music that they have laboriously reduced to notation’.³³ A transcription is indeed a ware, one that individual scholars can use, entrepreneurially, to manifest credibility and bolster claims over particular parcels of academic real estate. The functionality of transcription is disciplinary in a broader sense, too; transcription is one of the primary methods through which a particular brand of music studies – one that has its home within the specialist environments of academic music departments – is rendered distinct from the myriad other existing music studies which are thriving across the humanities, social sciences, and STEM fields. That is, the great portion of scholars of music who engage with transcription are employed by departments of music, where they are tethered to reproductive curricular structures that place enormous value on the techniques and pedagogies (ear training, sight singing, score analysis) required to create, read, and interpret documents containing Western staff notation. Within this disciplinary locus, transcription is an absolutely crucial tactic for evincing and performing a direct connection with sound.

In many regards, the practice of musical transcription is exactly what I’ve been describing here: it is a disciplinary ritual which involves the devoted labour of inscribing the audible into a visual form. For anyone who has engaged in trying to create a vivid visual representation of a musical or sonic event, the notion of transcription as an intense communion with sound should resonate. And for anyone who has included transcriptions in their research output, as product, the notion that transcription can be credentializing and also revealing of ‘ownership and control’ (as Nettl has it), should ring at least somewhat true.

³² Winkler, ‘Writing Ghost Notes’, 199.

³³ Nettl, ‘I Can’t Say a Thing until I’ve Seen the Score’, 87 (emphasis mine).

But, clearly, transcription is not simply a fervent, theological act; nor is it merely a disciplinarily directed practice, one whose primary horizon is the fulfilment of obligations to one's field or employer. To think that transcription is exclusively either of these things can only serve a mystificatory purpose.

One of the most telling moments in the 1964 Symposium is when Willard Rhodes discloses the minor revisions he made to his manuscript as he readied it for publication, calling attention to the deletion of 'one measure of music near the end of the piece which does not exist, but somehow found its way into my original transcription'.³⁴ If any passage in the Symposium puts ontological pressure on musical transcription, it is this one. Because the labour of transcription is often taken to involve, as Nettl suggests, a transcriber's emotional tie to music's sound, music that 'does not exist' upsets the communion between the transcriber and the music to be transcribed. But what is this thing to be transcribed if it is not something that does not exist? In a fit of humdrum metaphysics, Jaron Lanier has recently argued that '[w]hat makes something fully real is that it is impossible to represent it to completion'.³⁵ Applying Lanier's logic (which I find to be rather perverse), one might say that transcription – as representation (and let's not forget that representation is always both practice and product) – will be, in the final instance, defined by a lack, by an eternal inability to fully conjure up some full real thing that it indexes.

A more subtle way to think about representational ontology might be found in John Law's poignant book *After Method*. In it, he muses on the messy nature of representation, which he calls '*allegory that denies its character as allegory*'.³⁶ Rather than inhering in lack, allegory produces a surplus; it is, as Law puts it, 'the art of meaning something other and *more* than is being said' or 'the craft of making several things at once'.³⁷ If transcription is taken as allegory – and I think it should be – we might modulate the 'axiomatic perspectives' presented in the first pages of this introduction and emphasize how considerations of accuracy, apparatus, non-human presence, technique, iteration, repetition, perception, otology, audibility, measurement, inscriptional value, functionality, partiality, contingency, and context aren't necessarily encumbrances but themselves open up a vast field of potential performances and condition a politics that can be deeply sensitive and critical. Transcription is allegory, yes, but this does not mean it is some deficient other to the putatively fully present. Such a view would entail a rather fraught politics, one that would use a constrained understanding of *presence* and *representation* as the end point for assigning status and agency. That transcription will always be 'several things at once' is, perhaps, one of the reasons it has continued to exert such a gravitational form of attraction (and, sometimes, detraction) for at least certain segments of the academic music studies community. For sure, it is this multivalent sense of transcription that percolates most prominently through the six conversations below.

34 Rhodes, 'Transcription IV', 271.

35 Jaron Lanier, *You Are Not a Gadget: A Manifesto* (London: Penguin Books, 2010), 134.

36 John Law, *After Method: Mess in Social Science Research* (London: Routledge, 2004), 89 (emphasis in original).

37 Law, *After Method*, 88 (emphasis mine); 157.

The Forum

Between 27 September and 5 November 2013, I held six audio-only Skype conversations, each with two scholars who have utilized and theorized musical transcription in their research, with the simple goal of establishing a snapshot of current thinking about transcription amongst leading practitioners. These twelve scholars were not only strewn across a number of time zones; they came from a fairly wide variety of fields: ethnomusicology, popular music studies, linguistic anthropology, musicology, animal studies, music theory, philosophy, computer science. The decision to assemble this 'Forum' came close to this issue's copy deadline, meaning that time was limited, as was space in this journal's pages. Consequently, it was enormously difficult to decide whom to invite; almost immediately the list of scholars whose transcriptional work had piqued my interest over the past few years ballooned to a number that could not be accommodated.

In the end, I decided to speak with researchers whose work fell into six broad thematic areas: transcription in ethnomusicology (Tara Browner and Michael Tenzer); transcription of song lyrics (Dai Griffiths and Jennifer Roth-Gordon); transcription in popular music studies, with a particular emphasis on microrhythmic analysis (Anne Danielsen and Fernando Benadon); transcription of non-human animal vocalizations (David Rothenberg and Rachel Mundy); transcription practices within the culture industry (Kiri Miller and Sumanth Gopinath); and transcription and music information retrieval, with an emphasis on computational ethnomusicology (Emilia Gómez and Parag Chordia). These thematic areas are not meant to be comprehensive or definitive, and one might imagine other transcription-related topics that could provide ballast for substantive discussion: transcription and intellectual property; transcription on YouTube; transcription and the digitalization of early music archives; transcription and turntablism ... the list could go on. My hope is that the six conversations presented below will coagulate into some kind of larger whole, one that will prompt further discussion on musical transcription, which, as of 2014, does remain one important horizon for those seeking to represent and analyse musical performance.

Finally, two graduate students at the University of Oxford deserve my effusive thanks; without their labour, it would have been impossible to put this forum together so quickly. Joe Snape, with an ultra-sensitive ear for the nuances of spoken dialogue, was responsible for the initial transcriptions of all of the Skype conversations (to give some idea of the magnitude of his task, the final number of words he generated was in the vicinity of 55,000). He also compiled the voluminous data on the appearance of articles containing transcription in the journals *Ethnomusicology* and *Popular Music*. Maria Luiza Gatto put together a massive and pristinely formatted bibliography that helped guide me as I prepared to hold the conversations and write this introduction. She also used her significant knowledge of the data analysis and statistical software package STATA to produce the four graphs found above in Table 1 and Table 2. And, last but not least, there are the twelve participants: to each I give my heartfelt gratitude for the generosity they showed with their time and ideas. I learned much from our conversations.

Conversation 1: Tara Browner and Michael Tenzer³⁸

Jason Stanyek [JS]: Tara, could you talk a little about the process of preparing an edition of Native American pow-wow music for the *Music of the United States of America* (MUSA) series? What were the primary challenges in putting that edition together?³⁹

Tara Browner [TB]: It took a long time to do. But I wanted to produce something that – for me, as a pow-wow dancer – accurately represented what was going on with the music, especially time-wise. And I wanted everything to fit the way that it *really* fits. My hope was that somebody would just be able to look at it and see how the music works. So, it's *graphic* in various ways, it's a way of using Western notation as a kind of graphic notation. Many people say that Western notation is not adequate for non-Western music, but I actually don't believe that. I think that Western notation – especially in the rhythmic realm – is a tremendously powerful tool.

During the process of doing the edition, I found myself learning about the music and engaging with it in ways that I didn't anticipate. From an academic standpoint, I probably know pow-wow music better than anybody else – that is what I do. I've been dancing for years, and I can sing some of the repertoire too. While doing the transcription, I suddenly found myself learning things about the music and the songs that I hadn't really noticed before, and that's when I started to realize what transcription can do. It provides a way to engage with music with a kind of depth and intensity that, just listening to it, you don't get. [See Figure 2.]

JS: Was this the first time you produced such a large-scale transcription?

TB: I had done smaller ones, and I had already developed a way of transcribing the music that didn't use bar lines, so I could notate the independence of parts that characterizes pow-wow music.⁴⁰ But yes, I had never done anything quite like this before, and I'll be honest: I'm never going to do it again.

JS: Why is that? Was it so arduous that...

TB: It took a *long* time, it *was* arduous. And getting Finale to do what I needed it to do was so difficult because you're going against the grain of what it wants to do for you. The problem with Finale is that it's set up for people to do straight stuff with bar lines, and it does these things automatically, so you have to work hard to disable some of its particular functions. I have a grad student, Ben Harbert, who was really wonderful; he was able to take my ideas and develop a way to realize them with Finale. It was not easy.

38 The conversation with Tara Browner and Michael Tenzer took place over Skype on 27 September 2013.

39 Tara Browner, *Songs from 'A New Circle of Voices': The Sixteenth Annual Pow-Wow at UCLA* (Middleton, WI: A-R Editions, 2009).

40 See, for example, Tara Browner, *Heartbeat of the People: Music and Dance of the Northern Pow-Wow* (Urbana: University of Illinois Press, 2002).

The image shows a musical score for 'Round 3' of a song. It is organized into three main sections. The first section, labeled 'Round 3 A', features a 'Head Singer' part with a melodic line and lyrics: 'yah ah wa - ah wah-ay wah - ah wah-ah - ay'. Below this is a 'Group' part with lyrics: 'ya - ya - ya - ya - hoh'. The 'Drum' part consists of a series of eighth notes, and the 'Dance' part is represented by a sequence of down and up arrows. The second section features a 'Head Singer' part with a melodic line and lyrics: 'yah ay ah - ay - ay ah - ah - ay yah - ah - ay - ah'. The 'Group' part has lyrics: 'yah ay ah - ay - ay ah - ah - ay yah - ah - ay - ah'. The 'Drum' and 'Dance' parts continue with the same rhythmic and step patterns. The third section, labeled 'B', features a 'Group' part with lyrics: 'hay ya - ya - ya - ya - ya - ya - oh way hay ay - ay yay hay'. The 'Drum' and 'Dance' parts continue with the same rhythmic and step patterns.

Figure 2 Transcription of ‘Round 3’ of an Oglala Lakota contemporary-style straight song as performed in Los Angeles on 5 May 2001 by the Native Thunder Singers. The head singer returns for the embellished vocal incipit that begins every round. When the group returns, the main vocal emphasis is always a quaver to a semiquaver off from the drum beats, which gives the song a feeling of forward motion. This is graphically illustrated through the relationship between vocal scoring and drum scoring. Originally published in Tara Browner, ‘Song All Women’s Exhibition’, in *Songs from ‘A New Circle of Voices’: The Sixteenth Annual Pow-Wow at UCLA* (Middleton, WI: A-R Editions, 2009), 81. Used with permission of the author and A-R Editions.

JS: Tara, I do want to get into the details of your transcriptions: the bar-line issue, the transcription of dance steps – these are exactly the kinds of things I’d like to hear more about. But Michael, why don’t you give an initial overview of how transcription has figured into your research.

Michael Tenzer [MT]: Sometimes I do transcriptions recreationally, because I know of no other activity that makes me feel stronger as a musician. I’ve tasted many flavours: I’ve been a composer, I’ve been a performer, I’ve been a scholar. If you compose a piece, sometimes

you get it right and sometimes it's wrong; if you perform, you make mistakes or you have a bad night. But when you're transcribing you have this intense interaction with sound. When I've finished a transcription and I feel like I've nailed it, I have an extraordinary sense of satisfaction that I don't get from any other music-related practice. I feel that transcription is *honest* knowledge because it's *just* music with no critical or ideological overlay or anything like...

TB: Oh, come on!

MT: Of course, in every transcription there's interpretation and analysis; you have to make a lot of decisions. I grant that. But even so, it's just your direct engagement with sound.

I also want to make one more point: Harold Powers, whom I was privileged to meet a few times, told me: 'You know, when I was young I needed a lot of transcription'. And I suspect he said this for the same reasons we're talking about here: because it helps you to get to know the music. He continued: 'But as I got older, I didn't need it as much, because I wasn't talking about those sorts of things anymore'.

JS: So there's a generational issue to raise as well, perhaps?

MT: I feel that at a certain point in life you need to confront the details of music so you can really embody it and get to know how it works. But when you get older, you're talking about the bigger issues.

JS: Tara, did you want to say something about the critical or ideological dimensions of transcription?

TB: I have taught it in the past, at UCLA, but – in part because I have colleagues who pooh-poo it – it's really difficult to get across to students why it's important. What I tell my students is that when you transcribe, you're doing a couple of things: you're engaging with the music itself, and you're also making a roadmap for other people to follow. And in doing that, to a certain extent, a transcription is your interpretation. But it really shouldn't be just that; you want to get it down on the page in the way that Michael was just talking about. A transcription just has to be right; it's either accurate or it's slop.

MT: There are objections to that, though.

TB: The thing is, Michael, you're a composer, and I suspect we might approach transcription a little differently. I come to it with an extra tool in my arsenal: I have perfect relative pitch, which is a handy thing if you want to do transcriptions – it's not too hard for me to transcribe melodies. The other thing that is *really* important to me – and I think you can see it in the pow-wow edition – is being accurate with rhythm. Because there are so many transcriptions of American Indian music which are like, 'oh yeah, here's a drumbeat', and just stop. There are just a few little drumbeats at the bottom of the page. I think that Native music has been subject to the *worst* transcriptions of just about any kind of music. Because there's a sense of, 'oh it's simple, it's primitive', and actually it's not, because what you're doing is dealing with the relationship *between* the drumbeat and the voice. Most of the time, they don't line up; and they're not supposed to.

JS: Michael, how do you think about this issue of trying to represent rhythms that might not align with basic pulsations, that might fall between the cracks?

MT: For me, it's a real problem because that's where your subjectivity comes in. You need to write commentary on your transcription and explain the places where you're not sure what's happening. But it depends on the situation. Do you know this incredible CD-set that was put out in France about twenty years ago called *Les Voix du Monde*? For the past couple of years in my ethnomusicology seminar, I decided to make it a multi-year project with my students to transcribe the entire thing. I've divided it up among the students, who have had varying degrees of expertise. Many of the transcriptions are not great, but some of them really are. I did one of some Swiss yodelers who are yodeling to the pulsations of a cowbell. The cow is just walking along so the bell is not steady. The yodelers are also not necessarily entrained to that pulsation, so they're floating all over the place. I did my best to count and to measure, but I *know* that I've done some violence to that experience, to what was happening out there. Yet I learned an incredible amount: I learned their pitch language, I learned how their phrases vary in length, I learned that there's some polyphony and some counterpoint. But as far as the rhythmic representation, I don't know if I could ever figure that out. Well, unless I did what Simha Arom does: interview the people who make the music and let them confirm or disconfirm my perspectives through listening to the recording themselves.

JS: But isn't representing pitch an equally complex task as representing rhythm? Tara, you write about precisely this in your edition of pow-wow music. And, in that edition, you very explicitly add quartertones to your transcriptions of the vocal lines. So, regarding the notation of pitch, you made a pretty considered judgment about which direction to go.

TB: You have to keep something in mind: for that particular project I was trying to create something as close to a Western edition as I could, and I was transcribing one specific performance. So, I think that Michael is right, pitch is often flexible, and notating fine pitch details is perhaps not as important as representing rhythm and structure. But I can tell you, Native people *do* talk about music as being in tune or out of tune. So I suppose I would agree with him in terms of transcribing certain non-Western genres, that pitch is not as important. But with Native music you have to be true to the particular performance, and that *is* important to me. That said, pitch representation is probably the *weakest* part of Western notation. For me, Western notation's strength is what you can do *rhythmically*.

JS: You both deal with traditions that are sonically dense; there's a lot going on in the performances you transcribe. How does one hear into the interstices of performance? What are your techniques for listening into the crevices, so to speak? Some transcribers just provide reductive rhythmic outlines, or they just go after the vocal melody or a basic harmonic structure, but both of you are engaged in producing full scores, which is quite a task. So, I'm interested in how you think about the practice of hearing detail within dense textures and how you go about realizing the scores you create.

MT: There are a couple of dimensions to my answer. Some of my transcriptions are a combination of transcription and composition in the sense that I fill in some parts based on my knowledge of how the music goes. That is, even though I might not be able to actually *hear* all of the details that are present in a recording, I know they're there because of my familiarity with the music. [See Figure 3.] The more you know about what's supposed to be there, the more you can reliably portray it. But, that said, in gamelan there *is* a lot going on in the interstices, and at a certain level of attention to detail, you can only accurately transcribe such dense textures by using a very sophisticated recording setup, with separate microphones on each musician (as you did in your pagode article, Jason⁴¹). [See Figure 4.] But for me, that level of detail – say, with ‘microtiming’, which is fashionable now in music theory and, I guess, in ethnomusicology, too – has not been relevant to my scholarship, yet.

TB: This is an interesting issue. Most people, when they listen to pow-wow songs, only hear the repetitive cycles. So they just transcribe a cycle, and say: ‘here’s the song’. But I understand enough as a dancer to know that this is enormously reductive.

MT: Absolutely! The discovery that cyclic music is *not* repetitive – that it has many non-repetitive dimensions – was a truly important one. For me, an essential task of current ethnomusicology is to do further work in this domain, precisely because we take that kind of repetition for granted and don’t focus on the change that’s happening *within* repetitive cycles.

TB: That’s exactly what went on with the process of transcribing the pow-wow performance. I had to do a number of things just to represent how pow-wow works *in time*. I transcribed an entire session, so there’s a structural and temporal dimension to the whole edition that goes beyond the frame of the single song. I wanted it to be as if somebody were actually going to a pow-wow. Part of me wishes that I could have put the announcer banter in there too, but I ended up just sticking to the songs. As I was working through each of the songs, I realized that there was a larger structure to the event, an overarching structure in terms of where things like ‘honor beats’ were, and how people – toward the end of the penultimate round – did certain things to signal there was one more, final round coming up: all sorts of things that I had never noticed before. The single most important insight I had – one that nobody had articulated to me before and that I hadn’t quite understood until I did the whole edition – was that all of the changes in volume happen with the drum and *none* happen with the singers; the singers are always singing at the same volume. Those are things I didn’t know! So, there were so many aspects of this tradition that I – as a thoroughly experienced dancer, and someone who goes to a lot of pow-wows – learned only by doing a *complete* edition.

41 Jason Stanyek and Fabio Oliveira, ‘Nuances of Continual Variation in the Brazilian Pagode Song “Sorriso Aberto”’, in *Analytical and Cross-Cultural Studies in World Music*, ed. Michael Tenzer and John Roeder (New York: Oxford University Press, 2011), 98–146.

The image displays a comparative transcription of two modernized sacred Balinese compositions: *Tabuh Gari Pengawak* and *Lokarya Pengawak*. The score is organized into three systems, each containing three staves: surface melody, core melody, and drums.

- System 1 (Measures 8-32):**
 - Tabuh Gari Pengawak:** Surface melody starts at measure 8 with a *p* dynamic. Core melody is silent. Drums feature a complex rhythmic pattern with a *rit.* marking at the end. A bracketed section from measure 17 to 24 is labeled "drums: batu-batu improvi. (sim. throughout when bracketed)".
 - Lokarya Pengawak:** Surface melody starts at measure 8 with a *p* dynamic. Core melody is silent. Drums feature a complex rhythmic pattern with a *rit.* marking at the end.
- System 2 (Measures 33-64):**
 - Tabuh Gari Pengawak:** Surface melody continues with dynamics *f*, *p*, *f*, *f*, and *p*. Core melody is silent. Drums continue with complex patterns.
 - Lokarya Pengawak:** Surface melody continues with dynamics *p*, *f*, and *p*. Core melody is silent. Drums continue with complex patterns.
- System 3 (Measures 65-72):**
 - Tabuh Gari Pengawak:** Surface melody continues with a *p* dynamic. Core melody is silent. Drums continue with complex patterns.
 - Lokarya Pengawak:** Surface melody continues with a *p* dynamic. Core melody is silent. Drums continue with complex patterns.

Measure numbers (256, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64) are indicated below the staves. The text "1st Kempur" appears at the end of the first and second systems.

Figure 3 A comparative transcription of portions of two modernized sacred Balinese compositions, *Lokarya* and *Tabuh Gari*. Originally published in Michael Tenzer, 'Integrating Music: Personal and Global Transformations', in *Analytical and Cross-Cultural Studies in World Music*, ed. Michael Tenzer and John Roeder (New York: Oxford University Press, 2011), 366. Used with permission of the author and Oxford University Press.

9 $\text{♩} = 114$

voz

La _____ la iá la iá la iá _____ la iá la iá

violão de 6 cordas

violão de 7 cordas

cavaquinho

pandeiro levada

pandeiro partido-alto

banjo

repique de mão

tantã

surdo

The image shows a full score for the Brazilian samba-pagode song 'Sorriso Aberto'. It consists of ten staves, each representing a different instrument or the voice. The top staff is for the voice (voz), with lyrics 'La _____ la iá la iá la iá _____ la iá la iá'. The tempo is marked as quarter note = 114. The key signature has three flats (B-flat, E-flat, A-flat). The time signature is 2/4. The instruments are: violão de 6 cordas (6-string guitar), violão de 7 cordas (7-string guitar), cavaquinho (cavaquinho), pandeiro levada (levada pandeiro), pandeiro partido-alto (partido-alto pandeiro), banjo, repique de mão (repique de mão), tantã (tantã), and surdo (surdo). Each instrument part is written in a specific notation style appropriate to its instrument.

Figure 4 Full score of the Brazilian samba-pagode song ‘Sorriso Aberto’ as performed by members of Pagode da Tia Doca in January 2009 in Rio de Janeiro, Brazil. Transcription derived from multi-track recordings and discrete videos of individual performers. Transcription meant to be used with the audio-visual ‘virtual roda’ interface, available at <<https://www.music.ox.ac.uk/virtualroda/>> (accessed 15 February 2014). Originally published in Jason Stanyek and Fabio Oliveira, ‘Nuances of Continual Variation in the Brazilian Pagode Song “Sorriso Aberto”’, in *Analytical and Cross-Cultural Studies in World Music*, ed. Michael Tenzer and John Roeder (New York: Oxford University Press, 2011), 114. Used with permission of the authors and Oxford University Press.

MT: I totally agree with Tara: there is much about music that you can't learn until you write it down. That's why we have books, too.

JS: Tara, one of the things that is striking about your pow-wow scores is that they include transcriptions of the dance steps as well as those for the vocal and drum parts.

TB: People who come from outside of the pow-wow tradition do not realize that for Native people, the dancing, singing, and drumming constitute *one* performative act; you *can't* separate them out. I wanted to come up with a way to represent that. What happens is that the singing and the dancing actually *accompany* the drumbeat. The drumbeat is the central thing: *bum, bum, bum, bum*. Everything that we as dancers do, and everything that the singers do too, works in tandem with that: the drum is central. I wanted to create a score that would get that across to non-Native readers – what the dancers were doing with their *feet* and their *legs*. I couldn't transcribe the movements of *whole bodies*; once you get above the footwork level, it's much more improvisational, with each individual creating their own style. But I wanted to get the dance in there because it's part of the performance; it's part of the music in some way.

JS: Could we talk a little about how each of you go about producing transcriptions? What technologies and techniques do you use? What's the process from start to finish?

MT: First, I abhor the use of slow-down programmes and so forth because transcription is, for me, *ear-work*. The most important aspect of transcription is pedagogical. If you do slow-downs, or you use electronic transcribers, you're missing the point. I always tell my students to set themselves up with a nice, big piece of paper with wide staves so they can actually see the canvas of what they're working with.

If I assign my students to transcribe a track with which they have little familiarity, I always stress to them: 'Don't do anything without evidence. No guessing. If you don't have evidence for what you're going to represent, you'd better not represent it'. The same holds true when I'm working on a transcription of music with which I don't have much familiarity. I'll give a specific example. There's a wonderful duet by two BaAka children that Simha Arom recorded but which he never transcribed. I did a transcription of it, and it took me a full week because the voices float, and it's so hard to know what pulsation they're working with. And their pitch is extremely flexible, so you don't really know what kind of mode they're in. So after listening to it maybe a hundred times, I finally zeroed in on a motif that they sang *sometimes*. Then I tried to see if the places where I heard that motif were cyclically aligned, whether I could find a pulsation that would allow me to locate it at regular intervals. And very, very gradually, over the course of (for me) a very beautiful week, a structure started to come into focus. The two singers were constantly improvising this tune, and there were about twenty different cycles, and every cycle was different, and only sometimes would that motif come back. Gradually, I pieced it together, and then I had a transcription in front of my eyes in which I was able to see exactly the structure, because I'd lined up the vertical moments and found that they're actually building the structure from a very simple kind of *cantus firmus*. But I couldn't see that until I went from one bit of evidence to another, gradually filling in the rest of it. [See Figure 5.]

Mbola: Girls' Duet
 from *Musique des Pygmées Aka*, UNESCO
 (Rec. Simha Arom ca. 1971
 Transcr. MT)

The musical score is presented in a single system with 19 staves. The first staff is a prelude. Staves 4 and 5 are labeled 'Cycle 4 (:23\')' and 'Cycle 5'. The music is in a 2/4 time signature with a key signature of one flat. It features a variety of rhythmic patterns, including eighth and sixteenth notes, and rests. Fingerings (1-4) and breath marks (inhales) are indicated throughout the score.

Figure 5 'Mbola', sung by two BaAka girls. Transcribed by Michael Tenzer from the UNESCO CD *Musiques des Pygmées Aka* (original recording by Simha Arom, ca. 1971). Unpublished. Used with permission of the author.

JS: When you're doing this, Michael, over the course of the week, while you're sitting at a desk with headphones, rewinding and re-rewinding, are you using paper and pencil?

MT: Paper and pencil, yes. Another thing is to sing it yourself; that's really important. You've got to get it in your body. Then you can *feel* it.

JS: Tara, I was interested in the 'Apparatus' section at the end of *New Circle of Voices*. In it you lay out your techniques and technologies in detail: metronome, tuning fork, Finale, playback through MIDI. Does the 'Apparatus' provide an accurate picture of how you went about doing those transcriptions?

TB: Yes, for the most part. But, honestly, that 'Apparatus' is a required part of every MUSA edition. I don't even sit at a desk; I tend to sit on the bed and I have a keyboard next to me. I have a metronome because I'm constantly checking the time; if you see the edition you'll notice that there are a lot of little metronome time changes in it. I start with the tuning fork because I try to get into my head what's going on with pitches, and I also check what they tune on. Usually these would just be the starting pitches. Periodically, I'd have checkpoints in between to make sure my ears were right, especially when the pitch would shift. But other than that, I just sat there with a pencil and paper and the first thing I did was write out the rhythms, especially the main theme. And I'd do all of the rhythms of just one round, and sometimes that was just *hellacious!* After that I would go back and deal with working out the pitches. I also did what Michael just described: I'd learn the parts well enough so that I could sing them myself. I mean, you just have to get the music at that depth. But at the same time you want to be careful because sometimes you learn it and then you realize that, even though you think you know it so well, you're doing it wrong and you have the *wrong* version inside your head and you basically have to smack yourself to get it right. But it was a pretty low-tech process. Most of the time I didn't even use the MIDI playback until the end: I just wanted to *hear* what it sounded like once it was on paper.

It's hard work, just *grind-out labour!* It took forty, fifty hours just to be able to figure out the rhythms of that first round. I'd just work at it and go, 'no, this is not right!' and I'd have to go back. I tell my students that it's a discipline; it's character building, like learning how to play an instrument. It's hard to explain; I think it has to go from being work to being natural, and that's when you get to the point where you know you're good at it. There are many people who don't value it as much anymore but I do think it provides us with valuable roadmap. Through transcription you are communicating with an audience of people and you're trying to show them the things that you have discovered about this music. And that's why the quality of your work is so important.

JS: Michael, do you give transcription projects in your ethnomusicology seminars?

MT: Well, I downplay the whole concept of ethnomusicology; to me it's all just music, and I'm interested in teaching people to be the best musicians they can be, and to have rich musical experiences. I think it's a mistake to get into any kind of critical – or even anthropological – perspective on music until you're really solid in your musicality.

I teach transcription in *all* my classes, even in my first-year classes. And that goes for my non-music students as well. For non-majors, I teach a class called ‘Musical Rhythm and Musical Experience’. I have them do an assignment in which they have to represent mostly children songs and folksongs, songs that they can hear the number of beats in each phrase and how the phrases relate to each other. It’s not a note-by-note transcription, but it represents the form. If I were redesigning the curriculum, transcription would be a significant percentage of what students do through the undergraduate years. I have had some of the most eye-opening experiences trying to get people to transcribe Bach partitas, or get the rhythmic profile of a Ligeti *Étude*; these kinds of assignments do so much to cultivate musical sensitivity. And if it’s a Western piece, when you compare your transcription to the score, you learn a *lot*. To compare transcriptions with one another, and to see the different ways that people represent things – to me that is superb pedagogical material.

JS: It’s certainly true that the creation of transcriptions involves incredibly intense labour, as you said earlier, Tara. But *reading* transcriptions, and taking account of them in detail also involves hard work; I can take one page of transcription and spend *weeks* on it, trying to understand its intricacies. I’m thinking about my experience teaching graduate-level ‘Introduction to Ethnomusicology’ seminars and having the students read Paul Berliner’s *The Soul of Mbira*, with all of its incredible transcriptions.⁴² I’ve taught the book a zillion times and, to be perfectly honest, I don’t think a *single* student has ever taken the transcriptions in that book seriously.

MT: We should get down to a really important fact here – and this is what Ter Ellingson says in his article: the greatest value of transcription is for the transcriber.⁴³

JS: Yes, I agree with that, but the user of the transcription needs to be factored in too, I think. I don’t want to sound too cynical, but sometimes I think my students just don’t spend the requisite time with the transcriptions included in a given book or essay; they just pass right over the transcriptions, as if these were invisible.

MT: Yes, some people’s eyes glaze over when they see a transcription. But I know that not all of them do. Some people *do* get it. You can’t expect the world to be full of folks who love transcription! But another way to think of that quote from Ter Ellingson is not to call it transcription but to call it *representation of sound*. Because there are other technologies of representation – and we haven’t really touched on these yet. There are many things you can’t do with Western notation. You can’t show repetition very well (all you have are repeat signs which don’t really do the trick), you can’t show cyclicity, and so forth. There are many other kinds of graphic representations for these aspects of musical performance. There is now software that can help you do 3-D representations;⁴⁴ there are a lot of creative ways you can represent sound.

42 Paul Berliner, *The Soul of Mbira: Music and Traditions of the Shona People of Zimbabwe* (Chicago: University of Chicago Press, 1993 [1978]).

43 Ter Ellingson, ‘Transcription’ (see, for instance, 147).

44 For an example of such a 3-D representation, see Benadon’s transcription in Figure 10 below.

JS: Yes, but what's the difference between *visual representation* writ large – like a painter making a visual illustration of a piece of music – and *transcription*? Is transcription a subcategory of representation? Are there visual representations of sound that are not transcriptions?

MT: I don't know, that's a very good question. The first thing that comes to mind is something about representing sound as a temporal process.

JS: Yes, transcription needs to account for the temporality of sound, whereas we can think of a lot of different kinds of visual representations of sound that *don't* take account of its temporality. I would agree with you on that. Tara, do you have any thoughts on this issue, on the difference between transcription and other forms of visual representation?

TB: What it comes down to, for me, is that as a transcriber you're also a kind of *translator*. You're a person who is trying to communicate to other people. Michael is correct in that ultimately the person that gets the most out of the transcription is the person that's doing it. But you're also translating this for other people who have varying abilities of being able to take it in. The majority of the students I deal with on a daily basis don't read music and, for them, I create large-scale formal diagrams. You can pick up a lot of this stuff by using such diagrams in combination with *a lot* of listening. But in the end, for people who can read Western staff notation, following a transcription can give something vital. . . . I don't think you can replicate it any other way.

Conversation 2: Dai Griffiths and Jennifer Roth-Gordon⁴⁵

Jason Stanyek [JS]: You are both involved not only in *creating* transcriptions of words – of song texts and also of ethnographic and oral history interviews – but you've both also spent some time *thinking* about the practice of rendering sung or spoken texts into written ones. That's why I decided to invite you to participate in this forum on transcription. Dai, I'll start with a question for you: I recently read your article on 'Midnight Train to Georgia', and at the end of it you state that your 'contribution concerns the nature and activity of transcription', a very big claim.⁴⁶ What were the stakes in writing that article, and what are the views on transcription that you present in it?

Dai Griffiths [DG]: It goes back a long way, over ten years, to when I became interested in writing about the words to songs. I was interested both in music and in poetry and poetic theory and I always found that perspectives on words in songs tended to be quite reductive. And it seemed to me – this goes back to a transcription I did of an REM song for an article published in 2003 – that part of the problem was to be found in the way words in songs were presented as though they were 'page poems', with a hard left-hand margin, a free-

⁴⁵ The conversation with Griffiths and Roth-Gordon took place over Skype on 5 November 2013.

⁴⁶ Dai Griffiths, 'Words to Songs and the Internet: A Comparative Study of Transcriptions of Words to the Song "Midnight Train to Georgia", Recorded by Gladys Knight and the Pips in 1973', *Popular Music and Society* 36/2 (2013), 268.

flowing right-hand margin, and organized as lines.⁴⁷ For many years what I've done could be summed up like this: I put a hard-right margin, so that the words are represented according to the musical line. With the 'Midnight Train to Georgia' article I was interested in how song lyrics were faring on the Internet, where you would think there was a lot of room for experiment. But in fact, online transcriptions of lyrics all seemed to look very much as before: hard left-hand margin, as if they'd been transcribed from words straight onto the webpage. So my interest was to say, 'well, it's the "same old, same old" for transcription in the Internet world; here's a different way of representing lyrics that I've been developing for many years, and this might work on the Internet'. [See Figure 6.] I think that line about 'the nature and activity of transcription' was meant to say that the activity of transcribing is as interesting as the destination in some ways. One *learns* in doing the actual activity of transcribing.

Jennifer Roth-Gordon [JRG]: I guess I come at this from a different perspective, because as a linguistic anthropologist, that's where we start: *at* transcription. The core principle is that you're representing what was said or what was sung. You're taking that as your source of data, with a *lot* of attention to how it's performed, down to small details that include things like repetition and pauses. These details can be very significant, and we don't edit these things out because they're vital sources of information. So I transcribe lyrics, but I'm also very interested in how lyrics are quoted in conversation. I'm interested in lyrics as both sung *and* spoken, in how lyrics become so important to people in their daily lives that they actually invoke them when they're speaking. [See Figure 7.]

JS: Jen, could you speak a little bit about the transcription conventions you use in your work: indications for vowel elongations, increased amplitude, intonation, abrupt cut-offs, these kinds of things? For me, as someone involved in music studies, they seem decidedly sonic markings; they're very much like the notational symbols that we use in music scores. That is, they add a high degree of sonic detail to what are essentially visual representations. After all, transcriptions of the kind we're speaking of here are in many ways simply visual symbols on a page.

JRG: In terms of the linguistic analyses that I do, mine are far from being at the most intricate end of the spectrum. Really, there are many linguistic anthropologists and linguists who are much more driven than I am to put on the page as much detail as possible of what is heard and how it's performed. I'd say I go fairly light on that, and this is because I want my transcriptions to be very *readable*. Linguistic anthropologists are used to reading transcriptions. For lots of other people that level of detail is often difficult to absorb, because they're being forced to pay attention to how things *sound*, and you have to function on two levels: you have to pay attention to what was said – the content – and you have to pay attention to the form, which is what your attention is being drawn to through those transcription conventions. For my purposes, I'm much more interested in the actual con-

47 Dai Griffiths, 'From Lyric to Anti-Lyric: Analyzing the Words in Pop Song', in *Analyzing Popular Music*, ed. Allan F. Moore (Cambridge: Cambridge University Press, 2003), 56–59.

0.00 Instrumental introduction

									x	x
	x	x	x	x	x	x	x	x	x	x
GK:	x	x	x	x	x	x	x	x		x
										Mmmmmmmmm

0.13 Verse one

GK:	x	x	x	x	x	x	x	x	x	x
PIPS:	mmm		L	A						pro-----vcd

GK:	x	x	x	x	x	x	x	x	x	x
PIPS:	too	much	for	the	man				he	couldn't

GK:	x	x	x	x	x	x	x	x	x	x
PIPS:			so	he's	leaving	the	life	mmm		he's
	make	it								

GK:	x	x	x	x	x	x	x	x	x	x
PIPS:	come	to	know-----		Oooh	oooh	oooh		He	said he's

Figure 6 Transcription of the opening of ‘Midnight Train to Georgia’ as recorded by Gladys Night and the Pips. Originally published in Dai Griffiths, ‘Words to Songs and the Internet: A Comparative Study of Transcriptions of Words to the Song “Midnight Train to Georgia”, Recorded by Gladys Knight and the Pips in 1973’, *Popular Music and Society* 36/2 (2013), 246. Used with permission of the author and Taylor & Francis Ltd.

15	Babú: <i>Fala baixo. Não explana não, mané.</i>	Babú: Speak quietly. Don't spill it, fool.
16	Karate: <i>Assinou 155 [roubo], 157 [assalto a mão armada],</i>	Karate: He signed 155 [robbery], 157 [assault with a deadly weapon],
17	<i>121 [homicídio]. [Ele canta:] 'Mas se eu</i>	121 [homicide]. [He sings:] 'But if I
18	<i>tomo esse fulano, não tem pá, não</i>	take out this guy, there's no pá [sound],
19	<i>tem pum, eu vou ter que-'</i>	there's no pum [sound], I will have to-'
	<i>['assinar o 121.']</i>	['sign the 121.']
20	<i>Só com (som).</i>	Only with (the music).

Figure 7 In this excerpt, Brazilian *favela* (shantytown) youth joke about crime and police harassment, referencing Racionais MC's (1998) song 'Diário de um Detento' (Diary of an Inmate). Lines in bold are lines quoted directly from the song. Originally published in Jennifer Roth-Gordon, 'Linguistic Techniques of the Self: The Intertextual Language of Racial Empowerment in Politically Conscious Brazilian Hip Hop', *Language & Communication* 32/1 (2012), 40. Used with permission of the author.

versational flow, or the flow of a rap song, without getting too bogged down in levels of detail that have the benefit of giving you rich sonic information, but also often take away from readability, away from attention to the interactional dynamics, away from performance elements such as flow.

So, in terms of the lyrics of the songs I study, I transcribe those often just to form a sense of what the 'original' was so I can get a more precise understanding of the songs that are referenced in everyday speech. I work a lot with Bakhtin's idea that utterances are filled with other utterances, and shot through with other people's intentions. So, if someone quotes a rap lyric, they're picking it up from the context of a song and a particular musical genre. Inserting that into a conversation, turning it into a line of dialogue as if it were just spoken, maximizes this intertextual gap between the actual song and the conversation. I'm very interested in rap lyrics that have been decontextualized from their original location and then recontextualized. I spend a lot of time on the transcription of that act of recontextualization, when music is folded into daily conversation.

JS: Dai, in your 'Midnight Train to Georgia' article you mention at one point that that song is situated in the 'soul genre', and I was wondering if there are particular challenges to transcribing soul vocal performances.

DG: With that song, I had a basic issue of representing the Pips properly as well as Gladys Knight. If you remember, in some of the other transcriptions that I look at in the article, the device I used – simple parentheses for the Pips – really breaks down at some point. But I also had to decide on some quite technical issues of spelling words and, indeed, what counted as a word. The first word of the song is 'L.A.', the famous two letters, but actually before that there's a big 'Mmmmmmm', and putting that in place was important. My method has always been listening to the track dozens, if not hundreds, of times, and I've

always just used Microsoft Word as my basic software. Sometimes I've had to think hard about the *space* of a line. Often, I have to use a smaller font size to get the words in. So those kinds of quite technical issues are something I had to work out, and rather heuristically; I didn't bring any sort of theory to it. In each case I worked it out as I went along.

JRG: That's something I face a lot. In transcription, you're really jumping from something that's heard to something that's visually perceived. And once you do that, there are things you want to retain and convey even though they may not seem visually necessary; they're part of your connection to what was heard. So, for me, spacing is an incredibly important issue, because I'm also dealing with translations. I have to include the Brazilian Portuguese and this means that I have a lot more to fit on a line. I've had to fight with publishers over this – their idea of what should fit and my idea of how it needs to look are very different, as is the idea that a line *means* something, visually, that it serves to make a connection to what was heard, or indicates what should go together because of the way that it was performed. That's really something that I pay a lot of attention to. Publishers, on the other hand, are not as interested in these issues. They might break things up in ways that work for their space or layout concerns. But I want a passage to appear a certain way to convey a particular *sound* for people. This is something that I have had to stand my ground on.

JS: One interesting overlap between the work you both do involves taking account of colloquial language. What kinds of challenges do non-conventional forms of speech present to the transcriber?

JRG: You can think about such difficulties in terms of representation; once you put something down in writing, the conventions of standard language are overwhelming. Sometimes a word or a phrase that is well accepted in spoken language – something people wouldn't even pause on or recognize as non-standard at all – draws attention to itself when it's represented on the page. And then the way that people form a perception of the speaker becomes highly indexical. Or, I could say, 'perception of a vocalist', because this happens in rap lyrics all the time. So you have to be really careful about, on the one hand, wanting to be very true to what was heard and actually performed and, on the other hand, recognizing that it isn't a *true* or *accurate* or *equal* representation to just transcribe what was said.

DG: Yes, there is this issue around non-verbal vocal utterance, and colloquialisms, and I bring this up in my article on 'Midnight Train to Georgia'. I also seem to remember there's some writing by Dick Gaughan, the great Scottish folk singer, on the transcription of words, and this vexed him greatly, the way that Scottish folksongs had been rendered in more or less standard English.⁴⁸ In some repertories there is a great deal at stake in these issues, of course.

JRG: There are some interesting overlaps here with the issue of what I call *sound words*. Young Brazilian Portuguese speakers will often use words like *bum* and *pam*, and those are

⁴⁸ See for example Dick Gaughan's 'A Rant about Lyrics Sites' at <<http://www.dickgaughan.co.uk/songs/about-lyric-sites.html>> (accessed 28 January 2014).

sounds; they're recognized as sounds. You'll find them in comic books, for example. And yet Brazilian youth use them as part of their conversations, as placeholder-kinds of words, or to organize discourse, and the idea there is that they're very evocative; they *are* sounds, and the listener will hear them as sounds. Once they're written down though, they become invested with a kind of power and official status, and standard Portuguese speakers whom I interviewed about it were highly offended and confused and frustrated by this, because to them those were *not* official words, and so to transcribe them as if they were was very disturbing.

JS: You actually make use of your transcriptions in subsequent interviews, don't you Jen?

JRG: Yes, to get feedback, I play recorded conversations for either the speakers themselves or for people who were not present at the original conversation. Using transcriptions during these feedback sessions has become an important part of my methodology too. Holding a tangible piece of paper and looking at words printed on the page often messes with their understanding of what it is that they're hearing and focuses their attention on different things. That's when transcription takes on a special salience for my interlocutors, too.

JS: How do you deal with the intransigence that words sometimes have toward being transcribed? I know, Jen, that one of your convention markings is a question mark in parentheses, indicating that 'transcription is not possible'. I'm wondering about those moments when there is such a high degree of ambiguity or semantic noise – or literal, sonic noise – that the transcription becomes almost impossible to realize. How do you both deal with those moments?

DG: I'm very concerned with accuracy and rectitude – well, as much as I can be. I can think of an example I use in teaching from *The Fall*, where I transcribe the song 'It's a Curse' by Mark E. Smith, where I simply had to do my best with a vocalization that was very hard to render as words (I think I just put in 'ffffff'). Those kinds of moments have been pretty rare, but when they crop up, I simply do my best to be as accurate as I can.

JRG: Those situations are not rare for me at all and I think that, in some ways, I have a different view on accuracy. You *do* listen to these recordings over and over and over in the act of transcription, and you are interested in getting it as 'accurate' as possible. But you *can't* achieve total accuracy, especially when you're talking about spoken speech, and especially when you're transcribing a conversation in which lots of people are talking at once. It's interesting to think about these matters on a theoretical level; indeed, on two theoretical levels. The first is something we've already discussed: the sense that *accuracy*, getting everything down on paper, giving as much information as possible about how something was performed, isn't in fact *accurate*. This is because the transcription itself already changes people's perception, because they're *reading* all of this information about music or spoken speech. So, you're not in fact getting an accurate view of *how* something was performed, because you're being given instructions on how to *recreate* it. I think that's one level on which we can destabilize the idea of accuracy.

But it is the second level that I find even more theoretically interesting. Here I'm thinking of the mistakes I make in my linguistic interviews because I am *not* a fluent speaker of Brazilian Portuguese. There were times when I transcribed something incorrectly and this incorrect transcription spurred conversation and discussion about what I had transcribed versus what was said, and what was meant. This idea of accuracy – or *inaccuracy* – applies as much to people listening to music as it does to people having a conversation: you don't always hear what somebody has said 'accurately'. And sometimes this leads to very salient and overt forms of meta-linguistic discussion: 'Why did you say that?' 'No, no, that's not what I said.' Often, people are just working hard to figure out and make sense of what they heard, regardless of what was said. That's actually a really important part of communication. And if you think about it, it's an important part of people's perception of music too. It's a classic situation, the idea that people get words to a song wrong, and then debate over what the words are, or mumble parts because they don't know what the line is actually saying. Those misunderstandings or different perceptions are an active part of conversation, and an active part of music appreciation as well. People don't always get the words 100% right.

JS: Dai, do you want to say something more about that?

DG: In my work, an elephant in the room is the relation of the transcription to the track. It would be the same with trying to do a sheet music transcription. Rhythmic notation will only get you to a certain degree of accurate correspondence; something will fall beyond that. And that's how I approach the transcription of words. I'm doing my very best to pin the words down. But, on the microrhythmic level, the exact positioning of a word within the line *will* elude me. But of course I would assume that somebody is looking at the transcription while listening to the record. I mean, it would be very bizarre not to. So it raises this question: 'why are you doing this transcription in the first place?' As I said, we can only come back to this: it's part of a wider argument to do with poetic theory. I don't know who is listening to that argument, but I think the idea about 'verbal space', which I raised in one of the first things that I wrote in this area, has become established in a certain kind of popular music literature.⁴⁹ On the particular issue around the way that words look, it may be that I'm having something of a private conversation. We'll just have to see how that work develops in the future. What I did do – I think it was in the Cat Stevens article – was to review the way that other writers had done similar things to me in this particular kind of rock/pop repertory.⁵⁰ But, as I say, it remains to be seen if this work is going to be developed in the future.

JS: Jen, why do you do transcription?

⁴⁹ Griffiths, 'From Lyric to Anti-Lyric'.

⁵⁰ Dai Griffiths, 'Internal Rhyme in "The Boy with a Moon and Star on His Head", Cat Stevens, 1972', *Popular Music* 31/3 (2012), 383–400.

JRG: I think because, as a linguistic anthropologist, I'm interested in really understanding how language works, and how it is that we convey meaning through a language, and how that meaning is conveyed not just by the content of what we say, but through linguistic form and interactional dynamics. The whole process of meaning-making is highly relevant and interesting to me, and therefore the only way I can get at this is by looking at actual transcriptions of conversations, working through these, and trying to get as much of that meaning on paper for my readers to see as I can. Then I can then discuss it and analyse it and think about the significance of what it is that people are saying and how they are saying it.

DG: I would say that's why I'm doing transcription in the first place. Like any act of analysis, one always comes out noticing things that one hadn't spotted in the listening. I remember the 'Nightswimming' track of REM that I transcribed all those years ago: my transcription brought out something that no doubt I'd *heard* many times before, which was Michael Stipe shifting the place of the phrase 'nightswimming' in the line.⁵¹ I was quite charmed by that. And in the case of the Cat Stevens track 'The Boy with a Moon and Star on His Head', I knew there were internal rhymes there, that's why I chose it, but I'm sure that the activity of transcription and analysis brought out lots of subtleties in the placement of those internal rhymes that I simply wouldn't have noticed in the listening. I think John Covach might say, 'if you're going to screw something, you use a screwdriver'; the tool does determine a little of what you see. Or hear. I think that the activity of transcription will always bring its own consequence, as well as the point that you're trying to make.

Conversation 3: Anne Danielsen and Fernando Benadon⁵²

Jason Stanyek [JS]: In your scholarship, you each use what we might call modified versions of standard Western notation to represent what is very difficult to represent: groove, swing, microrhythm. Could you speak a little about some of the tactics you've developed to represent the rhythmic features of the music you study?

Anne Danielsen [AD]: Yes, I use standard Western notation and I think I've used it mostly to represent what I call 'rhythmic structure' or 'rhythmic figures', because much rhythm research relies on the premise that rhythm is an interaction between structuring schemes and various sounding versions of those schemes, which might be played in different ways. And I have tried to develop a way of representing those two levels of rhythm through notation. I use standard Western notation to represent the structural aspects of rhythm, and I add various arrows, circles and other symbols to highlight the microrhythmic design of a particular rhythmic figure. [See Figure 8.] This works quite well for some music. For example, James Brown's grooves are tightly structured and each layer of rhythm has its

51 Griffiths, 'From Lyric to Anti-Lyric', 56–9.

52 The conversation with Danielsen and Benadon took place over Skype on 14 October 2013.

The image shows a musical score for James Brown's 'Sex Machine' (1970). It consists of four staves: Vocal, Guitar, Bass, and Drums. The tempo is marked as $\text{♩} = 106$. The key signature has two flats (B-flat and E-flat), and the time signature is 4/4. A 4:3 time signature is indicated above the vocal line. The vocal line includes the lyrics 'Get up'. Circles and arrows are used to highlight counter-rhythmic patterns in the guitar and bass lines.

Figure 8 Counter-rhythmic pattern in James Brown's 'Sex Machine' (1970), indicated by circles. Arrows refer to early or late timing. Originally published in Anne Danielsen, *Presence and Pleasure: The Funk Grooves of James Brown and Parliament* (Middletown, CT: Wesleyan University Press, 2006), 77. Used with permission of the author and Wesleyan University Press.

own identity; it's easy to distinguish between the layers and identify the figures that are played by the instruments. In other grooves, however, in which the sound is more dense, it can be very difficult to distinguish between different rhythmic events. In those cases I would rely more on waveforms and sonograms, and other kinds of visual representations of sound. [See Figure 9.]

Fernando Benadon [FB]: The way I think of it is that standard notation is essentially an X-Y graph, a way of representing the location of pitches (Y) along a timeline (X). When you start to gradually modify the appearance of that graph (mainly by discarding the pitch, or Y component), you find yourself using 'graphical' representations. But these modifications simply provide a way to zero-in on one of the graph's two variables: the temporal (X) dimension. The degree to which you modify that graph depends, as Anne was saying, on the type of music that you are studying and on the properties that you want to highlight. For instance, Figure 10 shows 'aerial view' representations of time in two versions of Bach's C Major Prelude. Despite being in the same tempo, Richard Egarr's timing fluctuations are visibly more pronounced than Wanda Landowska's. The peaks at bar 23 indicate that both performers emphasize the arrival of the dominant pedal. Notice also how Egarr uses a consistent timing formula (a groove?): the two 'ridges' that run NW to SE in his graph reveal a recurring deceleration around beats 2 and 3 of each bar. In short, this type of visualization works well for large-scale sections made up of repetitive note sequences, whereas other graphing approaches would work better in other contexts. In Figure 11, for example, we see a mix of standard notation, microtiming annotations, and graphed data. While the top transcription is rhythmically accurate, the bottom transcription does a better job of conveying the phrase's rhythmic structure as a chain of malleable semiquavers.

AD: One thing Western notation is very good at is depicting the basic rhythmic structures of groove, especially in some forms of music. So, I really agree with Fernando that it's very effective for representing the order of events, or the basic structure of an event. Of course, it also depends on whom you'd like to communicate with; they have to be people who actually read music.

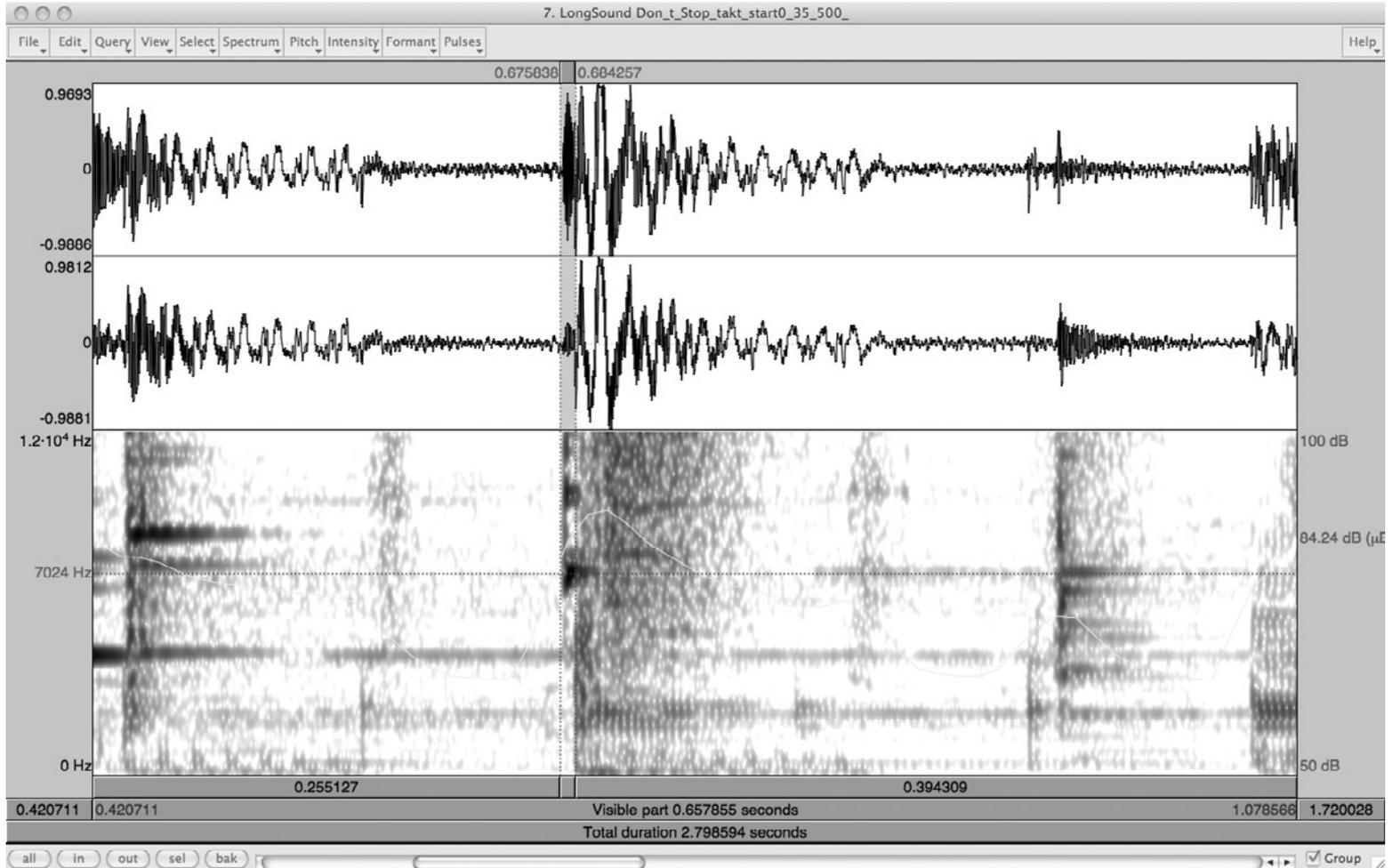


Figure 9 Sonogram of 0–12000 Hz (lower part), amplitude (upper part), and intensity graph (variable line in lower part) of detail showing bottle sound located approximately nine milliseconds (highlighted area) before the bass drum on beat 1 in the groove of Michael Jackson’s song ‘Don’t Stop ’Til You Get Enough’. Originally published in Anne Danielsen, ‘The Sound of Crossover: Micro-rhythm and Sonic Pleasure in Michael Jackson’s “Don’t Stop ’Til You Get Enough”’, *Popular Music and Society* 35/2 (2012), 157. Used with permission of the author and Taylor & Francis Ltd.

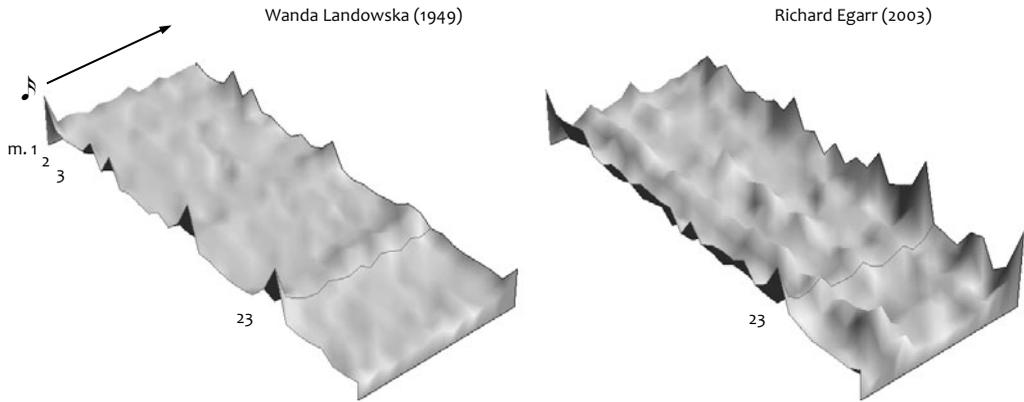


Figure 10 Contrasting timing profiles in two harpsichord renditions of J. S. Bach's C Major Prelude from *The Well-Tempered Clavier*, bb. 1–32. Each bar's semiquavers unfold as shown by the arrow. Terrain altitude corresponds to note duration – the more elongated the note, the higher the peak. The coordinate line marks the dominant pedal at b. 23. Transcription by Fernando Benadon using the computational software Mathematica. Unpublished. Used with permission of the author.

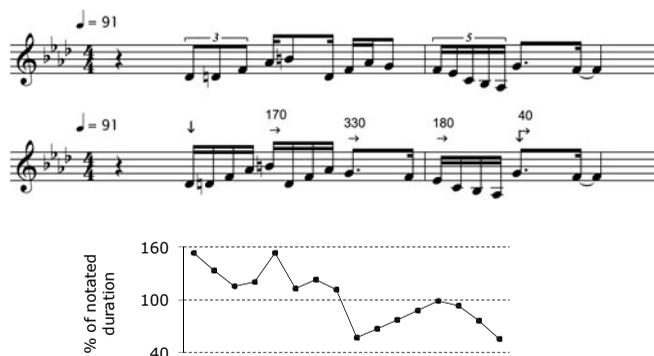


Figure 11 Coleman Hawkins (tenor saxophone), 'One Hour' (1:37). Two transcription versions of the same passage. The bottom transcription supports a hearing in which Hawkins' beats lag by 170, 330, 180, and 40 ms. This results from the semiquavers being first played 'too slow' and then 'too fast', as shown in the graph. Originally published in Fernando Benadon, 'Time Warps in Early Jazz', *Music Theory Spectrum* 31/1 (2009), 7. Used with permission of the author and the Society for Music Theory. Published by the University of California Press.

JS: So in that regard, what's the value of using these types of notations to represent groove and swing? I'm playing devil's advocate here, but in the case of analysing performances found on specific recordings, couldn't you just write prose descriptions that orient readers' ears to relevant passages on those recordings: this groove here, this anticipation there?

FB: You could, and sometimes we do. You say something like, 'this phrase is played 100ms behind the beat', and there's no need for a graph. However, it's important to say '100ms' because if the prose is imprecise – for example, 'here the playing is very relaxed' – you're leaving so many scenarios open to interpretation. You have to be precise about what you are describing, with or without a graph.

AD: I agree. I always combine notation or intensity graphs or waveform representations or sonograms with verbal descriptions of what's going on in the music. Both notation and other ways of representing rhythm absolutely need interpretation; they don't speak for themselves.

JS: Do you think that the way you each hear rhythm and groove is in some way dependent upon the practices of transcription you've engaged in? In other words, what comes first? Does the practice of transcription produce a different sensibility for understanding a rhythm? Or do those sensibilities produce the need for different transcription practices?

FB: What drives me to transcription is the awe that results from hearing something that leaves me puzzled. Only then do I worry about how I'm going to transcribe it. Once I begin that process, I might opt for one method over another, based on what I want to highlight, or because of mere convenience. But I hope that my listening habits are still 'pure' in the sense that I don't immediately bring the act of transcription into listening.

AD: I think I have become much more aware of microrhythms through transcription work. I also think that my listening has become more precise, in a way. I'm able to identify smaller microrhythmic deviations now than I actually was when I started. So yes, I think doing transcriptions – listening in this kind of way – has affected my approach to analysis. But also affected the way I actually *hear* these grooves. This is positive in many senses but it might also have a negative effect: I might hear a groove somewhat differently to an average fan, because I'm so focused on its microrhythmic design.

JS: In the most mundane sense, how do you each go about transcribing a particular musical performance?

FB: Some transcriptions are relatively straightforward. Usually, though, they're not. I've recently been working on a transcription of a Brad Mehldau unaccompanied piano solo. I know what the rhythms are, most of the time. And I know what the pitches are because I can spend hours on the piano figuring them out. But placing them on the right beat is *such* a difficult task, because often whole phrases are syncopated and displaced, and sometimes they're not. Since there's no reference – you don't have an accompaniment to tell you where the downbeat is – you have to infer a lot from the little expressive things he does. In other cases, you know exactly where the 'one' is, so that's never the problem. You may have to slow down the recording by 200% in order to figure out what's going on. Recently, I've been doing some tabla transcriptions, and those performances go by so fast! Unlike the Mehldau situation, the downbeat is never in doubt. Rather, it's the sheer speed of the performance that poses a challenge. For the first time in my life, I've had to transcribe sixty-fourth-note sextuplets! Of course, this is not something that you hear; it's a trill. But there is a very mechanistic process going on when the tabla is being played that way, so I feel it's important to be able to transcribe that. Obviously one has to slow down the recording considerably.

JS: What do you use to slow down the recording?

FB: Sometimes I use Audacity, sometimes Peak, sometimes Pro Tools. Whatever slows it down works for me.

AD: When I started, I used to slow down things using Logic. That was about eight or ten years ago. But today the situation is rather different, because there are so many music software programmes that can produce waveform representations. These are very useful for identifying the things that you're not able to grasp because they pass by so quickly. And of course, it's a particular problem when you deal with groove-based music, as I mostly do, because the basic pattern you want to transcribe only lasts for one or two bars – it's very short. And new software tools are *very* useful in order to kind of freeze time. You can map sound to the visual representation and figure out what kind of pattern it is, or where it is actually placed in time, or what kind of structure it has.

JS: Could you speak a little about how you go about using waveforms, as you do in your article on Michael Jackson's 'Don't Stop 'Til You Get Enough'?⁵³ What's your actual process, and what function does waveform analysis play as you attempt to disarticulate some pretty thick textures? Actually, that's one of the fascinating things about that particular article; you notate not only lead lines and harmonic progressions, but also rather dense fabrics of sound.

AD: Especially when the sonic fabric is so dense, as it is in this example, the waveform representations are very useful. I actually cut out a representative part of the groove, and then I look at it, for example, using PRAAT, which is a freeware program (it's the one I used for the Michael Jackson analysis). Then I take it apart layer by layer, and try to hear the individual figures. I try to work out the more standard transcription based on that, and then I start to measure where the actual rhythmic events are placed compared to that transcription. Is the cowbell played late or early? By how much? How many milliseconds? Does that answer your question?

JS: Yes, it's a great answer. Forgive me if some of these questions seem obvious. I find these technical issues really fascinating – how people go about doing the work that they do; how technologies, at least in part, constitute transcription practices. Are we in a different realm right now in 2013 than people were twenty years ago, or fifty years ago?

FB: It's not *all* that different from say, 1999, or even before then. We're still measuring manually, as Anne was describing. Even though we're getting to the point where we can have the computer do that by itself with various algorithms, the onset detection problem is still a very difficult one to automate. Let alone interpretation – that obviously still falls on us. I do find myself spending, regrettably, way too much time just measuring stuff. Then again, it's reasonable to think of onset measurement as part of the transcription process.

JS: Fernando, how do you do this measuring? Do you work with waveforms?

⁵³ Anne Danielsen, 'The Sound of Crossover: Micro-rhythm and Sonic Pleasure in Michael Jackson's "Don't Stop 'Til You Get Enough"', *Popular Music and Society* 35/2 (2012), 157.

FB: I either look at spectrograms or look at waveforms, whichever gives me the most information. I mainly use my ears in the end, because appearances can be deceiving, and you want to try to combine what you're seeing with what you're hearing. I'm not really too much of a 'techie' in that regard. If I can figure out where the onset happens, and if I can do that with a fairly crude approach, then it all works out.

AD: The ears are really crucial, I think. I use waveform representations and sonograms but if we don't have the sound, it's pretty difficult to figure out what's going on. So it has to be combined with sound. Of course, when you get accustomed to working with waveforms and sonograms, you can get more out of them without listening. But I still think the ears are absolutely crucial.

I want to point one more thing about the 'onset problem' that Fernando just mentioned. Sometimes, the software detects onsets where no onsets are present. In these cases, you have to look deeper into the music and the representation to find the *relevant* onsets. And moreover, it's a problem if the music does not *have* clear onsets. In much music there are very clear onsets; if there's a percussive sound, it's much easier to have a good grasp of where the onsets are. But if it's a very dense or muddy type of sound, it can be extremely difficult to decide where to place the onsets. But perhaps the notion of an onset as a particular point in time is not very relevant. That's actually one of the things I'm working on at the moment: how can we deal with extended onsets?

JS: You both work, as far as I know, with already recorded music. You don't actually go and do recordings yourselves, do you? I'm asking because for a project I did on Brazilian samba-pagode, the only way I could do the transcriptions that I felt were necessary for the research, was by bringing all of the musicians into the studio, recording them on Pro Tools while also videotaping them, and going back to the US with these discrete tracks that I could actually *look at* – see the musician's hand positions, for example, or locate which part of the instrument is used to produce which particular sound. You both work on musics that can sometimes have extremely dense textures, stubborn textures that are resistant to transcription. My solution with samba-pagode was to go to Brazil and to record everybody separately. But that's not possible in all cases. So, I'm wondering if there are moments when you can't really hear what you need to hear. If so, what do you do?

FB: That happens all the time! I mean, it's not an exact science, as much as we want it to be. Therefore, you give yourself as much licence as you think is ethically acceptable. While you would not want to be inserting onsets if you have no idea where they should go, a bit of educated guessing is often required. It's just how it is: a lot of textures are very thick and the recordings are not always very good. Still, typically the stuff that I'm interested in foregrounding will have to be drawn from a passage where I am confident about the location of the onsets. I need to know that the data are reliable.

AD: While there are millions of tracks where it's very, very difficult to hear what exactly is going on, it helps when you're working with music in a style that you're familiar with; in those cases, you know what to listen for. I've listened to a lot of Afro-Cuban music – and that's of course related to things I've done with funk and with African American music –

but because I don't know what the stylistic conventions are, at least not to the same level as, for example, funk, I find it difficult to be able to identify the right things. Of course, this is something we're confronted with all the time.

FB: Having the video, as you said Jason, can be so helpful. Lately I've been transcribing a lot of drum solos, and I learned very early on that I should just do the ones that I have video for. Not so much from the perspective of microtiming, but just to know *what happens when* – the timbres get so mingled together, and everything happens so quickly that it's important to have that video cue to help differentiate between, say, the high and middle tom-toms, or between the bass drum and floor tom. These types of distinctions can only be made – at least to my ears – with the help of video footage.

AD: And, as I mentioned earlier, I also use PRAAT, or other software tools that allow you to isolate or separate different sounds down to the 10ms level, or even smaller; then you can hear 'is this actually a snare drum, or is it a percussion instrument, or a combination?' By way of the software, you can say 'OK, it's actually two instruments here'.

JS: What's the *value* of doing transcription of the sort you're both describing, this micro-level transcription that's extremely nuanced and attempts to access the inner workings of musical textures?

AD: I think those tiny small differences in timbre and timing, in duration, in phrasing – all these tiny small differences – are actually crucial in order to grapple with this thing we call 'musical style'. Sometimes you find the same rhythmic figure in two styles, and on paper you might mark it as being exactly the same. But you can detect vast stylistic differences by way of those small, tiny differences. Such nuanced transcriptions are absolutely crucial in order to identify systematic variation and distinguish between musical styles.

FB: Yes, I'm 100% in agreement here. What I'm trying to show is why a particular music is special, in a concrete way. Even though some people think we're nuts to be focusing on such tiny nuances of a performance, these are not as tiny and nuanced as one thinks! And that's another challenge. At what point do you just chalk it up to noise and say: 'Well, I did measure this one little thing here, but I don't think anybody on the planet can hear that.' On the other hand, you might say: 'That was a very fine-grained temporal effect, but I definitely heard it and I hear it every time.' The distinction between those two ends is actually very subtle – a few milliseconds away from each other – and learning to distinguish between them can be tricky. As long as I'm focusing on the appreciable aspects of the music, then I feel perfectly fine with what I'm doing.

Conversation 4: David Rothenberg and Rachel Mundy⁵⁴

Jason Stanyek [JS]: Clearly, non-human animals vocalize and sing and make song. And, clearly, humans have been fascinated by these vocalizations for a long time. But historically,

⁵⁴ The conversation with Rothenberg and Mundy took place over Skype on 15 October 2013.

what have been some of the reasons behind the urge to create visualizations of these non-human animal vocalizations? Why transcribe these at all?

David Rothenberg [DR]: The simplest answer is that nobody would take these things seriously as scientific data until they appeared on paper. Sound comes and goes too quickly. So there was no science of the study of animal sounds until they could be written down or diagrammed in one way or another. It's the same with a lot of data in science, actually; sound had to be turned into an image before it could be studied by anyone. Before there was recording of sound, any manner of visual approaches were taken to write down these sounds that seemed to have organization and structure. Sometimes they sounded close enough to musical notation for musical notation to be used, but other times these sounds were nothing that could be encompassed in musical notation. So people made up imaginary words or diagrams or pictures, and before the advent of recording in the late nineteenth century there's a whole history of different kinds of naturalists – particularly ornithologists – trying to figure out how to notate these things one way or another. The entire field really got much more scientific when spectrograms were invented, which were for a whole different purpose as you know; they were invented to teach the deaf how to talk. And it turned out that spectrograms were a great way to represent sound that couldn't be depicted very clearly with existing notations.

Rachel Mundy [RM]: I think the question of *why* people wanted to transcribe these sounds has changed a lot over time. There are transcriptions of animal sounds by Athanasius Kircher, a Jesuit from the 1600s. And he was transcribing things like sloths. [See Figure 12.] In these early cases, I think that music notation was a useful way to convey sonic information to people who couldn't hear the original sound. But I agree with David that, in the late 1800s, there was a shift in thinking about visual information. Personally, I tend to blame it on laboratories at the end of the nineteenth century wanting to quantify information visually through graphs of all kinds of things: graphs of pulses, graphs of respiration. And I suspect that that's maybe one of the starting points for trying to understand why graphs of animal sounds seemed more important than talking about animal sounds as music. In some ways, it might be very similar to why transcription was important in the study of non-Western music.

DR: One reason it's important is that these sounds aren't just *like* music, they actually *are* music. The reason I think a lot of animal vocalizations are music is that they are performances with a structure and a form; a beginning, a middle and an end, that are repeated over and over and over again, and there's a right and a wrong way to perform them. And the meaning is not some kind of meaning you can translate. They're not like a language where a specific message is getting across; it's something you have to hear many, many times. The way music is, you generally don't get bored hearing the same song over and over again, even though you got the message: you want to hear it again and again. That's how music works and I think that's how the songs of birds work. That's why people throughout history in many different cultures have called these sounds 'songs'.



Figure 12 ‘Musica haut siue pigritie animalis americani’ and ‘Figura Animalis haut’. Originally published in Athanasius Kircher, *Musurgia universalis sive ars magna consoni et dissoni in X. libros digesta* (Romae: Ex typographia Haeredum Francisci. Corbelletti, 1650), 27.

RM: Yes, in twentieth-century science there was this huge debate about whether to use music notation or a graphic notation of some kind like the spectrogram to represent animal sounds. And it’s very clear that one of the things that influenced the decision to use graphic notation was the urge to isolate animal sounds from musical enquiry. People talked about this, about whether or not music transcriptions were advantageous to studies in the sciences because they allowed you to talk about musical questions, like phrasing and the way that musical notation works. It’s geared towards musicality and issues in sound that are very specific to musical concepts. And so the question was ‘Should we have a transcription that makes those musical concepts that we believe we have in animal sounds *clear* to the listener, or should we have a notation that *excises* them, so you can be more “objective”?’

DR: I completely agree, but I would add that a sonogram is also a musical transcription. Scientists chose it because they thought it was ‘more objective’, but actually it’s another way of imposing human categories on sound. And if you look at how these sonograms are used, like in so much other scientific visualization, scientists *massage* the image so that it starts to show something that looks meaningful. You can’t just print out a sonogram – you have to adjust the factors; otherwise, there’s a lot of noise there.⁵⁵

⁵⁵ See a relevant discussion of the ‘spectrum analyser’ in David Brackett, *Interpreting Popular Music* (Cambridge: Cambridge University Press, 1995), 27.

RM: Exactly.

JS: Rachel, do you agree that sonograms are transcriptions?

RM: Absolutely, I think they are transcriptions. I'm just not sure they're musical transcriptions per se. Western musical notation, and in fact many other kinds of notation, use things like metre and phrase indications or tablatures to show patterns, gestures, and forms that are important to musicians. Sonograms and spectrograms, and the melogram in ethnomusicology, are very good at describing timbre and pitch contours – the elements Charles Seeger would have called 'descriptive' – without showing the more conceptual things we call melodies, phrases, or harmonies very well.

DR: Yes, you're right. But I think sonograms make everything look like music. They make speech look like music; they make the noise of a jackhammer or a cicada look like music. A sonogram of noise is more beautiful than a sonogram of a continuous tone. And what's great about a sonogram is that it allows things to make sense that, otherwise, might not make sense. In a way, a sonogram is better for noises than for pitches. Sonograms really make noise come alive as something full of structure and form, and really – even though they don't look like standard Western musical notation – sonograms somehow look like musical notation to me; they reveal hidden music in noise.

JS: David, you sent me one of your recent articles, 'The Investigation of Musicality in Birdsong', and in the email accompanying the article, you remarked that it was hard to get your co-authors to put even a small amount of music notation into the article and that, originally, there was much more than there was at the end.⁵⁶

DR: Yes, that's right. The whole history begins when I wrote *Why Birds Sing*⁵⁷ and I said: 'You scientists are not asking the most interesting questions about birdsong. You're not asking what makes the best song *the best song*; how these female nightingales are choosing some song over another. Not because it has the most notes, or the most complexity; it's something much more subtle than that. But you're not asking that question, because you don't think it's objective enough to measure. So you're denying the music that's there, and instead you want to say "more notes means more mating success"; you're just saying that because it's easy, there's no way that it's true.' And they responded: 'Why are you criticizing this? You don't know anything about science.' To which I would say: 'Well, do you think I'm *wrong*?' and they would avoid it. Finally, one scientist, Ofer Tchernichovski, said: 'You're absolutely right, we should do what you're talking about.' So four or five years ago he created this research project and had a number of post-docs trying to look at musicality in birdsong. It's very hard to get these people to really look at the musical side of things. So we finally did publish this paper, and even though it's co-written with all these other people who are neuroscientists and biologists, they insisted that I be the lead author, because the whole thing was my idea. I laughed about that because it's exactly the kind of writing I don't

56 David Rothenberg, et al., 'Investigation of Musicality in Birdsong', *Hearing Research* 308 (2014), 71–83.

57 David Rothenberg, *Why Birds Sing: A Journey Through the Mystery of Bird Song* (New York: Basic Books, 2005).

like – so detailed, very technical, and full of references. On the other hand, now that it's out and certain scientists are reading it, they say: 'Oh I guess you really are onto something serious here.' Because it's this kind of paper, they take it seriously. But if you write a *book* full of musical notation, anecdotes, historical tales and stories, and other 'trappings of the humanities', they just say: 'That's nonsense, that's not serious.'

RM: It's funny you used this phrase, David, the 'trappings of humanities'. I've been thinking about this a lot lately. Part of me feels that this issue of animal song transcription is neat because it lands right at the place where the humanities and the sciences have developed extraordinarily distinctive specializations in using evidence. And musical evidence operates very differently from evidence in an ornithological laboratory – I sometimes wonder if the problem lies exactly in what you're saying: in the 'trappings of the humanities'. Perhaps if we could figure out a way to share a common language about our evidentiary research, then we could bridge that gap a little bit more effectively.

DR: So, can you think of an especially musicological, ethnomusicological, or music-theoretical approach that nobody has yet tried to apply to the study of birdsong?

RM: I don't think that this is the answer that you want, David, but because I am who I am, I want you to tell me exactly what kind of bird we're talking about here.

DR: OK, like nightingales. I'm here in Berlin and Berlin is full of scientists studying nightingales. They have a long history of analysing what the birds are doing. And when you read their papers on the structure of the nightingale's song, it sounds much more like a software manual than anything musical. You hear the song and you think: 'Oh, this is so cool – they're going from this phrase into this phrase, and so on. How come nobody is studying this in a rigorously musical way?' And you know, the nightingale is one of the ideal birds to work on. Mockingbirds, too. Not many people are really studying the music of these birds. Hollis Taylor is one scholar who has done significant research on butcherbirds, and she's obsessed with transcribing. She has a hundred-page transcription of this bird's song; this is a bird that really sounds like someone whistling, has very clear notes, so it's very suited to musical notation – almost uniquely so in the world of birds. She's one example of someone really doing huge amounts of transcription. But that's just one scholar. I have a feeling that there are others out there who might have interesting things to say about this.

JS: What can ethnomusicologists, popular music scholars, musicologists, music theorists, *learn* from the histories of transcription that you both study?

RM: One really interesting place to go with that is to start reimagining the relationship between the humanities and the sciences in ways that music is particularly well-suited for, because people who do musical research tend to be interested in scientific questions as well as questions that we normally call humanities questions. My sense is that scientists are very, very good at *mapping what they know*. One place that the post-humanities is leading us is towards *a better understanding of what we don't know*, and one thing that it seems like we really don't know is what to do with species boundaries, how to productively use or think

about those boundaries. For me, music is a place that is an incredibly powerful entry point into that line of enquiry.

DR: Are you talking about species boundaries between different animal species, or something else? Is that what you meant?

RM: Different people have very different ideas about what a species boundary is. If you ask a modern geneticist, they'll give you one answer; if you ask a naturalist, they might give you a second set of answers. I think if you ask a music student, they might have a very different understanding of what 'species' is. So part of the problem is that we talk about species as though we all have a common understanding of that concept. But one of the things the post-humanities brings to bear is that we don't really have a common understanding; these are cultural concepts as well as a scientific ones. Perhaps more importantly, this dichotomy between science and culture doesn't always serve us very well, particularly when we talk about music. The idea of species is one example of this, but musical transcription – the topic of this forum – is another one.

DR: I'd like to return to your previous question, Jason: 'Why is it important to learn about the histories of transcribing animal songs?' I would say that it's important for anyone in musicology to confront a sound from a species that we can't really interview or get to know or talk to. How can you make sense of this phenomenon using the training from your discipline? What can you do with the song of a bird? How can you improvise with this musician from another species with whom you cannot talk? Can you cross species lines with that? Sure, there's a fairly clear boundary species-wise between birds and humans, but in terms of music, maybe not. After all, we often play music with people with whom we cannot speak, because they come from other cultures, and we do not share a spoken language.

RM: David, something that's nice about what you're saying is the idea that there are a lot of people who seem genuinely to have treated those animals as intellectual and creative colleagues. People like Messiaen, who everyone knows for his incredible birdsong transcriptions. I was joking with a colleague that we should call Messiaen a 'magpie', because he's a bird plagiarist. If birds were humans, we would accuse him of theft I think, looking at the way he composes.

JS: Speaking of birdsong transcriptions, Rachel, I was looking at your 2009 essay 'Birdsong and the Image of Evolution' from *Society and Animals*, in which you include a transcription of an American robin song – the cheer-up, cheerily, cheerio song – that is combined with a sonogram of the same song.⁵⁸ [See Figure 13.] How do you go about the business of actually transcribing birdsongs?

RM: I don't do it very often! I don't have the musical chops – it's very difficult. Some birds are easier – robins are not so terrible, but they have a sound at the end of some phrases that

58 Rachel Mundy, 'Birdsong and the Image of Evolution', *Society and Animals* 17/3 (2009), 210.

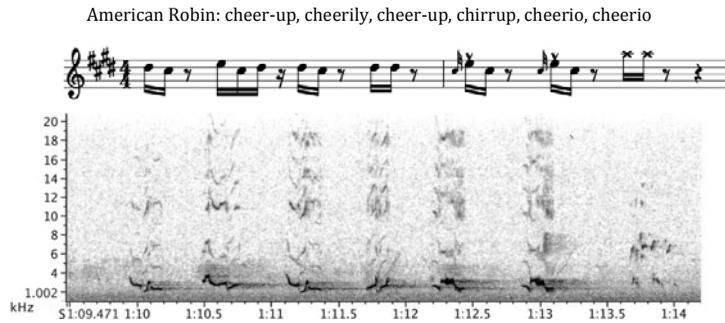


Figure 13 Transcriptions of an American Robin: *top*, mnemonics; *centre*, musical transcription; *bottom*, spectrogram. Revised from version originally published in Rachel Mundy, 'Birdsong and the Image of Evolution', *Society and Animals* 17/3 (2009), 210. Used with the permission of the author.

Donald Kroodsma calls a 'hisselly'.⁵⁹ If you don't transcribe the hisselly, it's not so bad with robins. But one thing David was saying is that animals have a tremendous timbral variety, and usually the real problem is when you hit a sound that's not like a straight sine wave in the way a flute sound or a violin sound is; in these instances of croaking or trilling or cracking or hollering sounds, the spectrogram really hits its stride, I think, because transcribing timbre is something that it does really well.

DR: One project I'm working on is to try and create a new way to represent some of these songs that combines the familiarity and tonal accuracy of traditional musical notation along with the kind of precision at representing noisy sounds that sonograms can be good at. So we're trying to tackle the song of the humpback whale, which is the longest, most drawn-out musical animal sound we know about. And it really hasn't been represented visually any better than the first paper about it from 1971 by Roger Payne and Scott McVay, in the journal *Science*.⁶⁰ No one has done a better graphical representation than the very simple hand-drawn tracings of sonograms they did then. So now I'm working with Michael Deal, a data visualizer, to try and come up with the *coolest* way to represent the song, and then we're going to try to publish it somewhere and make some interactive online version of it. He's working on this way of combining sonograms with musical notation. In terms of my own transcriptions, what I do spend a lot of time doing is adjusting visually the appearance of sonograms to make them clearly reveal the structure of a sound that I think is noteworthy. The most interesting ones are of a live performance I did on clarinet with a whale; it's clear that the whale is actually paying attention to me and responding. When you look at the sonogram, you can see that the whale is actually trying to sing a continuous, straight note of a single pitch, unlike his usual approach. You can *visually see* his efforts in the image, trying to keep a steady tone. The visual representation is quite helpful in illustrating the whale's response. [See Figure 14.]

59 Donald E. Kroodsma, *The Singing Life of Birds: The Art and Science of Listening to Birdsong* (Boston: Houghton Mifflin, 2005).

60 Roger S. Payne and Scott McVay, 'Songs of the Humpback Whales', *Science* 173 (August 1971), 587–97.

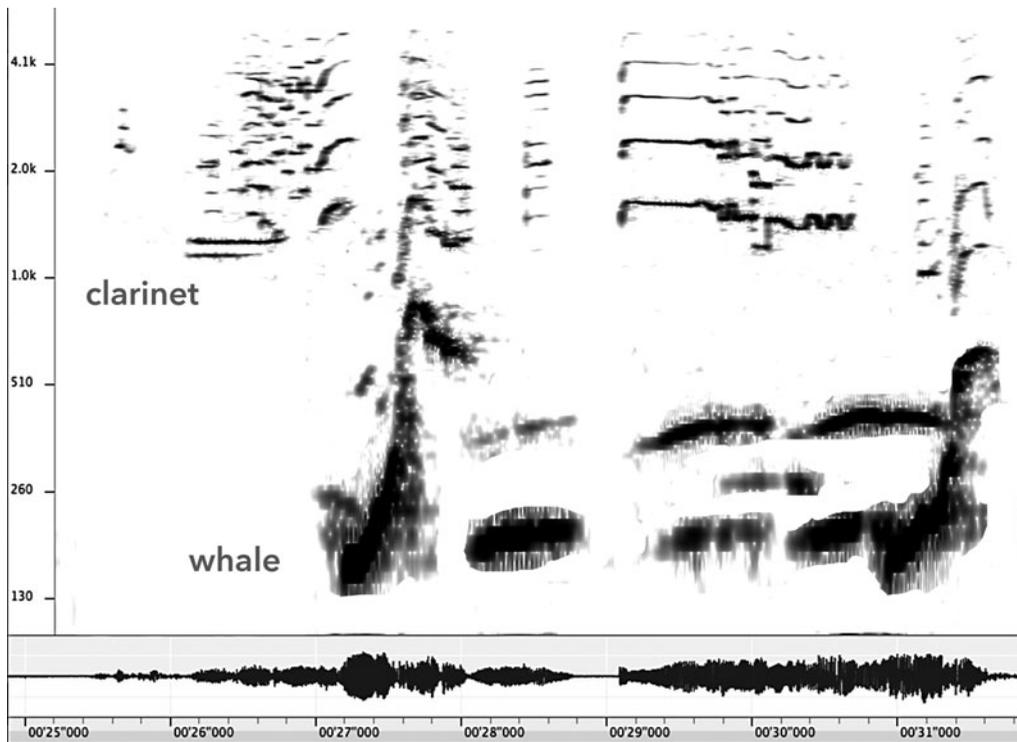


Figure 14 Sonogram of an excerpt from a live, unedited duet between David Rothenberg (clarinet) on a boat and a humpback whale underwater, with the clarinet sound broadcast underwater and recorded together with the singing whale via a hydrophone. After the clarinet plays a glissando up to 831Hz (Ab5) at 26'' the whale clearly responds with a high cry immediately afterwards at 26.5''. At 29'' the clarinet plays a steady tone that then becomes warbly and whale-like, after which the whale whoops again at 31''. From David Rothenberg, *Whale Music* (Terra Nova Music, 2008), track 2. Used with permission of the author.

RM: Well, why is that? What is the difference between that and hearing the response on a recording?

DR: People don't trust their understanding of sound as it comes and goes by. We take *visual* information much more seriously as data. We think it's something you can analyse, you can sit and stare at it rather than listen to it go by again and again. And I know some people are sonifying data to reveal its patterns and rhythms and things like that. But more often it goes the other way. Sonic information is turned into something visual, and then it is taken more seriously. That's something about the way human beings tend to use visual and auditory information.

JS: Rachel, do you agree with that?

RM: Let me put it this way: I was trained in analysing visual data and texts since I was a kindergartner. And the only place I've ever studied analysing sound is in music classes. So I do think that we don't spend a lot of time talking about how to understand sound and how to listen carefully and responsibly. Even if we're more capable of listening well than

we think we are, most people don't have sufficient background in the practice of critical listening. And the stakes of listening seem very high to me, since our capacity to listen, and in the case of transcription, share and responsibly interpret that experience, governs our ability to hear meaningful sounds in the world around us.

Conversation 5: Kiri Miller and Sumanth Gopinath⁶¹

Jason Stanyek [JS]: The reason I wanted to talk to you as a pair is that I see some overlaps between, Kiri, your research on *Rock Band* and *Guitar Hero*, and Sumanth, your research on ringtones.⁶² In the worlds that you've both studied, transcription figures as a form of labour that is harnessed by capital; you've both written about this. In the ringtone and videogame industries, how does transcription work?

Kiri Miller [KM]: Transcription plays a fundamental role in *Guitar Hero* and *Rock Band*, and these are some of the very best-selling videogame franchises ever published: millions and millions of people play them. The game developer, Harmonix Music Systems, would license popular music tracks, and then company staff would laboriously transcribe versions of, initially, the guitar parts, but then as the game developed, also bass parts and drum parts, into a special notation system that the players read in the course of reproducing the song during gameplay. [See Figures 15 and 16.] An enormous part of this transcription process is not automated; there's a significant amount of individual labour that goes into placing these 'gems' – which are the equivalent of noteheads – onto the notation track for each song. There are four different versions for each song – four different difficulty levels – so the easiest one might almost look like a Schenker graph, and then at the most difficult level there's almost a note-for-note correspondence to the original. To give a sense of the scale of the labour involved, over the life of the *Rock Band* series, the Harmonix audio team 'authored' over 23 million gems (that's how they refer to the transcription process).

Sumanth Gopinath [SG]: In the context of the ringtone industry, particularly with its two major phases – first monophonic and then later polyphonic – a similar sort of process took place. It wasn't as complicated as what Kiri just described in terms of transcription detail, but the amount was probably correspondingly greater in terms of the number of songs and musical tracks that were being transcribed. With the monophonic ringtone, typically the melody or some basic hook line of a popular song would be the source for the transcription. In the case of a polyphonic ringtone, the transcription would take particular vocal and instrumental parts in order to create a kind of simulation or a MIDI arrangement of a song. Transcribers would make use of some kind of Digital Audio Workstation, relatively standardized software, and essentially just listen to these thirty-second clips and transcribe the keyboard

⁶¹ The conversation with Miller and Gopinath took place over Skype on 16 October 2013.

⁶² For example see Kiri Miller, *Playing Along: Digital Games, YouTube, and Virtual Performance* (New York: Oxford University Press, 2012), and Sumanth Gopinath, *The Ringtone Dialectic: Economy and Cultural Form* (Cambridge, MA: MIT Press, 2013).

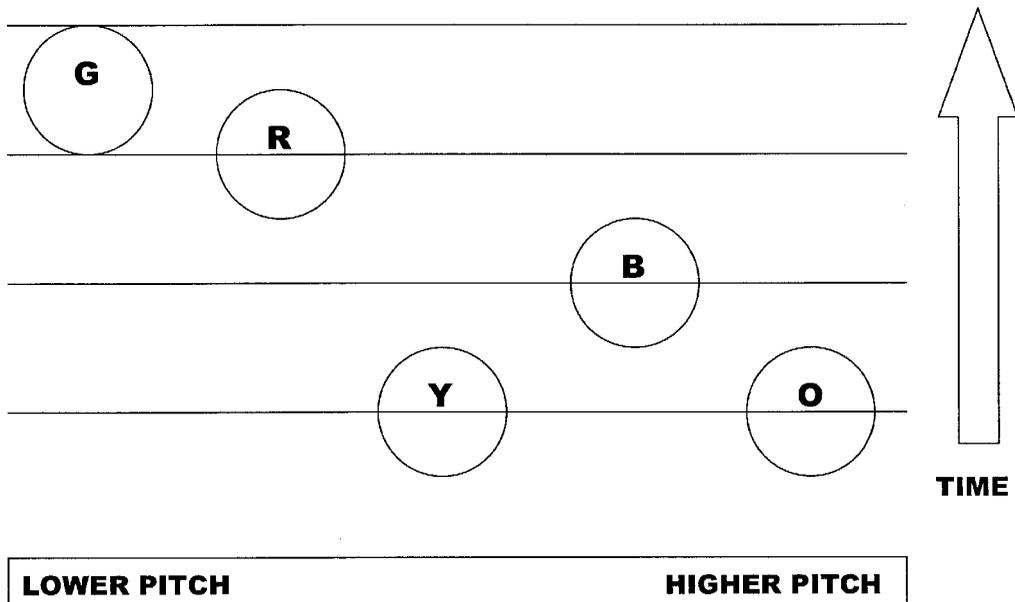


Figure 15 Schematic diagram of *Guitar Hero* and *Rock Band* guitar/bass notation. The games display the notes shown in the diagram falling from the top of the page to the bottom, but with perspective applied so that they appear to be coming directly at the viewer, like the roadway in a driving game. The letters in the diagram indicate the colour of the note (green, red, yellow, blue, orange). These coloured notes mirror the layout of the five coloured fret buttons on the guitar controller. As each note – or several notes, in the case of power chords – crosses a fixed reference line of coloured notes at the bottom of the screen, the player must fret and strum. In the games, sustained notes are represented with a long tail following a note head. Smaller note heads designate ‘hammer-on’ notes, which may be played by fretting without strumming. Diagram created by Kiri Miller. Figure and explanatory text originally published in Kiri Miller, ‘Schizophrenic Performance: *Guitar Hero*, *Rock Band*, and *Virtual Virtuosity*’, *Journal of the Society for American Music* 3/4 (2009), 400. Used by permission of the author and The Society for American Music. Published by Cambridge University Press.

part or the guitar part or the drum track. If they were going to use or adapt some aspect of the vocal line, they’d do that as well. As with *Rock Band* or *Guitar Hero*, this was not automated labour: a labour-intensive practice was involved.

JS: So in the cases you’ve both mentioned, were these transcriptions first done in standard Western notation, or were they directly made in the *Guitar Hero* and *Rock Band* notation system, or using MIDI?

KM: Definitely not first transcribed into Western notation – I can’t imagine what purpose that could have possibly served in this process. It was a combination of working by ear and using MIDI tools to create transcriptions in their own notation system that channel the ‘feel’ of the recorded tracks in the course of gameplay.

JS: I’m bringing this up, Kiri, because I remember reading in your book that Harmonix [the developer of *Rock Band* and the original developer of *Guitar Hero*] would always tout their employees as musicians.



Figure 16 Rock Band notation (screen capture by Kiri Miller). Originally published in Kiri Miller, *Playing Along: Digital Games, YouTube, and Virtual Performance* (New York: Oxford University Press, 2012), 91. Used with permission of the author and Oxford University Press.

KM: Yes, they do have a tremendous number of musicians, although I can't tell you how many of those musicians read Western notation; they're predominantly popular musicians who are accustomed to learning and playing by ear, although some, I'm sure, also would have had formal training with Western staff notation.

JS: So when they did the transcriptions they would be sitting with the original recordings, and then basically translate what they heard into this new notation system?

KM: Well, as far as the original recordings are concerned, they would have stems. So they wouldn't have to be listening to the full mix. But yes, basically this was a process of abstracting by ear, aided by waveform graphs and MIDI score tools. Also, many songs have multiple guitar parts, and they'd have to determine the 'Ur-guitar part' they'd want to represent in the notation. One of the key issues for these games is, of course, *playability*. I'd be really interested to hear if there's any analogue to that in Sumanth's case. These transcriptions were made for players to have a satisfying interactive musical experience. In this regard, Harmonix would certainly hear back from a lot of players if it were a *bad* transcription, if it didn't meet the expectations and needs of the players.

SG: Let me answer Jason's question, and then I'll answer your question, Kiri. I don't have as much information on the monophonic phase but I'm guessing that transcribers basically listened by ear, and then coded directly into the ringtone-text-transfer language (RTTTL), or the equivalent – there's also the iMelody format, and there are numerous other text-based formats. All of those systems are essentially text-based versions of staff notation. So

you identify note name, duration, register, and volume, and different text-based languages have different degrees of flexibility and capacity to represent a particular sound. In those cases I think they probably just went directly into that notation system. In the case of the polyphonic MIDI-based transcriptions, at least with regard to the examples I found out about, they generally used the piano-roll editor that was at that time common in Cubase and other Digital Audio Workstations. Transcriptions were done *visually* on a track-style layout, generating quantized or grid-arranged beats on which you could, with a cursor, plot or tap in what notes and durations were being attacked and held, etc.

With regard to issues of accuracy and player experience: obviously, with ringtones, the issue of playability didn't come up. But the issue of the quality and the effect of the ringtone was a big deal. For example, I spoke to a ringtone producer who was working for *The Source* – the hip-hop magazine/marketing entity – and they were creating a mobile ringtone channel that was like an app that you put on your mobile phone through the Wireless Application Protocol. They received a number of hip-hop ringtones that had been done in Japan, and a lot of the production was not usable because the Japanese ringtone transcribers would transcribe voices as if they were vocal melody lines. So you'd get these monotonous melodic lines that made the ringtones sound awkward and ineffective. *The Source* ended up giving these tracks to largely US-based producers and told them *not* to do it that way.

JS: Kiri, what are the feedback mechanisms for Harmonix to learn users' views on the quality of the transcriptions?

KM: They have very active online player forums. But before you even reach that point, there's a *lot* of play-testing in-house. So there's a long process of ensuring that a song feels right, and that it's playable and that the difficulty levels are appropriately calibrated. And that's a *huge* amount of work that any game company does: in-house quality assurance, basically.

JS: One important point is that with *Rock Band* and *Guitar Hero*, players or users are accessing the transcriptions directly; in a way, the transcription is the fundamental interface. Whereas with ringtones that's not typically the way it works; the transcription manifests sonically, but the user is not in most cases engaging directly with it (although there are exceptions, of course, and Sumanth, you write about these). Kiri, is there a relationship, say, between the graphical elements of the *Guitar Hero/Rock Band* transcription system and graphical notations that have been developed to augment or displace Western notation? For example, you write about the notation system of *Rock Band* and *Guitar Hero* as a *distinctly different* type of system. Was this system developed because of certain needs that Western notation – as a kind of standard system – couldn't fulfil?

KM: Well, this notation system really does have a different purpose. In the case of ringtones, it's clear that this new kind of transcription – the transcription code – is still going to be used to generate the musical output. Whereas the transcription in *Guitar Hero* and *Rock Band* actually no longer has any direct relationship to generating musical sound. Instead, what it's doing is giving the player physical instructions for physical performance, and so

the kinds of elements that it has to represent are the elements that help the player bridge the gap between what they're hearing and what kind of input they can give to their guitar controller. In some ways this notation is much like neumatic notation: it's all about melodic contour; pitch is relative – there are these little connected phrases and figures. There's no need for you to say 'this is an A and it's one notch away from a B'; that's not part of the transcription at all. The way that rhythm is depicted is also quite relativistic.

JS: Is it accurate to say that the system is 'prescriptive', to use Charles Seeger's formulation? That it indicates to performers how to create sounds rather than providing a description of what those sounds are?

KM: Well, it falls in a really funny place in the prescriptive/descriptive understanding of transcription. Because it's highly *descriptive* in the sense that it is an abstraction and a representation of a particular performance; that's the performance that's going to come out of the speakers, the sonic performance. And it's *prescriptive* in terms of the player's physical action.

JS: In your book, you liken it to four-hand piano transcriptions.

KM: I think there are a lot of parallels there. In the case of four-hand piano transcriptions, the transcription doesn't need to tell you how the orchestra plays it; it needs to tell you how to play it on the piano. And that's also the case here; the transcription is telling you how to play it on the guitar controller, and not how to play it on the electric guitar.

SG: So, if you could imagine a piano that would generate those kinds of orchestral sounds, in their full splendour, that would be actually quite a lot closer to *Guitar Hero* or *Rock Band*. In the case of the piano transcription, at least in some instances – we know that lots of people wouldn't necessarily have had access to or even have heard the orchestral or full ensemble version of something – a kind of *imagined* sound would have been a key part of the performance of such a transcription.

KM: That's right. And part of the promise of *Rock Band* or *Guitar Hero* is that you *get* the recorded sound; it's not *only* an imagined sound, although there are also dimensions of imagination and interpretation and role playing that may go into that.

SG: What you and Jason are saying is completely right, but, with the ringtone at least, the access to the notation is really not the point. Instead, it becomes a story about sound *fidelity*, at least that's how it's typically discussed in the industry. Hence, the evolutionary story of the progression of file format types that ends up pushing towards the sound file. At that point you're basically just working with the musical file that's already been created, and the only encoding system is the one that's part of the digital sound file recording system. There's no notation or transcription other than that.

JS: That's a huge shift, isn't it?

SG: Yes. As I talk about in the book, there are obvious economic implications; it changes the nature of the labour: from the highly skilled labour that's involved with polyphonic ringtone transcriptions, into a de-skilled form that consumers can do (although skilled

musicians are still typically hired to make sound-file ringtones for sale). This de-skilled form uses the one mode of representation that's not transcription into a kind of notation system or quasi-notation system like you find with *Rock Band*, *Guitar Hero*, or the Digital Audio Workstation's piano roll editor. Instead, the main representation is a kind of end in itself: the iconic depiction of the waveform. And that's something that users are now familiar with. There are all sorts of free ringtone programmes that use waveform graphs. Audacity is free, and people use it to create ringtones all the time.

JS: Kiri, I wanted to ask you a question about skill in the *Guitar Hero/Rock Band* industry. Are there transcribers in those companies who became known as particularly virtuosic transcribers? Ones with better ears? With a better ability to re-render an original tune in a form, as you mentioned earlier, that would be satisfying to the players of the game?

KM: I would assume yes, it's clearly a specific skill set, and some people are maybe more in tune with what feels more playable. But I haven't heard people talk about this very often, as far as citing individuals by name. There definitely was competition in terms of comparing different renditions of individual songs at the level of *industry*, because once Harmonix lost control of the *Guitar Hero* franchise and it was being produced by Activision, then *Rock Band* and *Guitar Hero 3* were in direct competition, and did have some of the same repertoire with different transcriptions. So at the level of company property, there certainly was quite a bit of discussion about which versions of a track were more musical, more playable, more fun, more difficult, or required more virtuosity. And there were partisans on both sides. So there'd be talk among users about two versions of a song with a really difficult guitar solo: which one was the better transcription?

SG: It was different with ringtones. You had many companies that were competing against one another, producing different versions of the same ringtone. One person I talked to, who worked for a company called 3rd Ear based in the UK, told me that within the company there was a real sense of who the best or most prodigious ringtone programmers were. He described one person as a 'ringtone animal', someone who would churn out great ringtones very quickly and was known for this in the recording industry press. Of course, it mattered whether you were doing it in a piecework, casual-labour way, or you were getting paid as a full-time, salaried employee. And I think this was one of the reasons why one of the big companies I looked at ended up paying their workers such high salaries; they wanted to set a certain quality standard that would look and sound *better* than the ringtones that other companies were producing.

JS: As we've discussed, both of you do research on musical practices that are heavily reliant on transcription. But do you make your own transcriptions to bolster your arguments about these practices?

KM: In my own work, I typically do not make any Western staff notation transcriptions. But what I often do is dwell on whatever forms of notation are in use in communities I'm researching. My work before videogames was on Sacred Harp singing, a participatory practice

5thSymphony:d=16,o=5,b=100:g,g,g,4d#,4p,f,f,f,4d,4p,g,g,g,d#,
g#,g#,g#,g,d#6,d#6,d#6,4c6,8p,g,g,g,d,g#,g#,g#,g,f6,f6,f6,4d6,8
p,g6,g6,f6,4d#6,8p,g6,g6,f6,4d#6



Figure 17 The top half of the figure shows the Ring Tone Text Transfer Language (RTTTL) instruction string for a monophonic ringtone arrangement of the opening of Beethoven's Symphony No. 5. The bottom half shows a realization of the RTTTL instruction in standard Western staff notation. The example reveals a monophonic condensation of an originally polyphonic texture with staggered entries in the string parts. The example also demonstrates numerous notational irregularities (no flat signs, semiquavers instead of quavers) and reveals rhythmic alterations to the underlying duple metre (particularly the rests). Reproduction of Figure 2.2 in Sumanth Gopinath, *The Ringtone Dialectic: Economy and Cultural Form* (MIT Press, 2013), 69. Used with permission of the author and MIT Press.

that's built around a very specific kind of notation system, a 'shape-note' system.⁶³ For me, in both cases, it seemed less relevant to transcribe what people were doing into Western staff notation than to take seriously and analyse the notation system that they were using. Recently, I've begun to look more carefully at dance games [*Dance Central* and *Just Dance*] and started to wonder more and more about whether motion capture is a form of transcription and, if it is, in what sense. I've also been thinking about the relationship between choreography and transcription and transduction, and how the issues we've been discussing intersect in a more multisensory, embodied system.

SG: In my ringtone book I include a second-order transcription of a Beethoven Fifth Symphony ringtone. [See Figure 17.] Here, you see this odd version of the piece because it's a literal transcription into Western notation of what you would find in the ringtone text-file format. And you can see that a couple of differences pop out. There are no bar lines, there is no metre, and all sorts of other things disappear. One is the difference between sharps and flats; these don't matter because this is essentially a twelve-tone system. Then you also find out that there are musical decisions that have to be written into the notation, such as extra pauses that the synthesizer reads literally to produce a particular musical effect. But second-orderedness is only one aspect to my transcription process. The other one is an attempt to capture and create an object to analyse. A lot of people use transcriptions in this way to study popular music or various other kinds of music or practices. Transcriptions do figure heavily in my other work (for example, in my research on Steve Reich) and there they allow me to do close readings of particular musical or sonic utterances, to peer into the implicit or latent aspects of a certain recording or performance.⁶⁴

63 Kiri Miller, *Traveling Home: Sacred Harp Singing and American Pluralism* (Urbana: University of Illinois Press, 2008).

64 Sumanth Gopinath, 'Reich in Blackface: *Oh Dem Watermelons* and Radical Minstrelsy in the 1960s', *Journal of the Society for American Music* 5/2 (2011), 139–93.

KM: Staff notation can definitely help you to represent certain kinds of questions and answers, though I guess it hasn't served that purpose in my own work. I'm concerned about audience and access, too. And I know that the folks at Harmonix made a big decision at the beginning that there would never be a 'little note' in their games, you know, a little musical note icon with a flag on it, because its mere presence would be intimidating or wouldn't seem fun or whatever. And, as we know, this is also often the case with music-oriented publications – many editors don't like to see a lot of notes in there because that might mean that nobody outside the discipline will ever read the book. I hope that that's not completely true, but I do worry about that.

Conversation 6: Emilia Gómez and Parag Chordia⁶⁵

Jason Stanyek [JS]: What is music information retrieval (MIR)? What are its primary innovations, its primary concerns and challenges?

Emilia Gómez [EG]: MIR was born as an interdisciplinary research field.⁶⁶ And its main goal – as its name says – is the *retrieving* of information. It often takes account of large collections of musical data and provides a means to retrieve music within those data fields based on certain descriptions or parameters. But I think the field has evolved in a way that means we cannot only restrict it to mere retrieval; it now has more to do with the processing of information about music. We can retrieve, but we can also transform, we can generate. Maybe Parag wants to add something?

Parag Chordia [PG]: I look at music as information. So if you consider all of the fields that come from computer methods, then this one is no different. My research is about how humans listen and whether we can model that with computers. The applications are wide-ranging and can be anything from corpus-based analysis to actually informing what we do in terms of music cognition and music perception. Until recently the humanities and music studies in particular have not been enamoured with computational methods, but of late, just as with other fields, there's been a turn, and we've seen that we can actually do really interesting things. We can start to provide answers to hard questions: 'What does it mean to listen?' and 'How can we do this automatically?' So some of what MIR does is very application-driven, and some is more theoretically driven; but the core of what we're trying to do is to model the process of listening, and obviously the computer can do this on a large scale.

Another thing is that music information retrieval really took off when we stopped thinking about retrieval as a pure transcription problem. It's an interdisciplinary research area: some researchers come from music theory, some from ethnomusicology, many come from

⁶⁵ The conversation with Gómez and Chordia took place over Skype on 15 October 2013.

⁶⁶ Two key publications on MIR and 'automatic transcription' are Anssi Klapuri and Manuel Davy, eds, *Signal Processing Methods for Music Transcription* (New York: Springer, 2006) and Zbigniew W. Rás and Alija A. Wiczorkowska, eds, *Advances in Music Information Retrieval* (Berlin: Springer, 2010). See also the articles on 'computational ethnomusicology' in the special issue of the *Journal of New Music Research* (42/2; 2013).

straight engineering, be it electrical engineering or computer science. And we all bring our different perspectives. The earlier strains – such as at Stanford’s CCRMA [Center for Computer Research in Music and Acoustics] – involved a lot of composers and music theorists. For them, much of it *was* a transcription problem: there is a sound, and they wanted to transcribe it into Western-style notation. And, actually, that is a really tough problem; transcription is a very specific learned skill that very few can do well.

EG: And of course, transcription is also a practice that is interpretative; each person can have a different way of transcribing any given piece of music.

PC: Exactly. So, the problems are much more fundamental than transcription. We really started to make progress in the field when we began to look less at the specific task of transcribing notes – which is difficult and raises a host of issues – and started to break down the broader research problems into smaller component parts: Can we get a computer to tap to a beat? Can we figure out what key a performance is in? What scales are being used? Do they conform to known tunings, or are they doing something different? What about continuous pitch motion? What about timbre? Once we looked at these types of questions we started, in my opinion, to take a much more user-centric approach, one that concentrated on the common features found in the act of music listening. This is in distinction to a standard music-theoretical approach which is, ‘OK, we want to generate Western music notation for all these different things so we can use our normal analytical tools’.

EG: It’s also because the music information retrieval field has been partially driven by some commercial applications, by the practical need to deal with large collections of music. One example can be found in ‘audio fingerprinting’, intended to identify a given audio sample in a database. Another one is ‘music search or recommendation’, based on computing *similarity*. And it’s because of those types of applications that people are addressing more fundamental research problems, trying to extract semantic descriptions from music content. In addition, these commercial applications focus on mainstream popular music; MIR research has become wedded to very practical uses and repertoire.

JS: Can you describe what audio fingerprinting is?

EG: Audio fingerprinting is essentially a process that analyses audio content in the service of song identification. This is, for instance, used in copyright management, when someone wants to know which radio stations are using a given piece of music or a given recording. Shazam, SoundHound or Vericast are fairly well-known commercial applications that use audio fingerprinting, which, in essence, is really about identifying the signal – the music signal or the audio signal – but not trying to describe it. Just trying to identify it – ‘it’s *this* piece of music’.

PC: I think it’s important to contextualize what we’re talking about. The big question is: ‘Why are we doing what we’re doing, and how does this really relate to unpacking a piece of music, understanding its cultural context and so on and so forth?’ To answer certain questions, you have to look at an entire corpus of music. So if you want to make a claim

like ‘melodies tend to move in small intervals’, that’s not a claim about an individual piece of music, it’s a claim about a large set of music. In order to examine these types of claims, you almost inherently need to do computational work. For instance, if you’re searching for either culturally specific universals, or cross-cultural universals, you need to analyse large datasets, and there’s no way to do that without getting computers involved. So that’s one of the really big motivators. A lot of times, claims are made that are very specific to a few hand-selected pieces of music, and these particular instances attain a status of generality. But what we really wanted to do was to say: ‘OK, if this is a really general claim, then we should be able to show this over the whole corpus of whatever it is that it is being claimed for.’ For me, that’s a really big motivator: MIR keeps us honest, and allows us to do large-scale, cross-cultural studies.

EG: I also think the fact that you have to actually *train* a computer how to do a certain task means that you have to formalize the way we would do it manually. That’s also something that is interesting about the use of technology; you can use it as a way to formalize expert knowledge or everyday knowledge. The formalization process involves concrete comparison of how different people would analyse, describe, or transcribe a musical piece.

PC: Yes, in many instances you’re better off using a computer as a tool. Let me give an example. In Hindustani classical music we’ve been debating for two thousand years how many notes there are in the scale. Well, there’s a sense in which that’s not a really well-formed question, because a note is not exactly a note, but this is *exactly* the kind of question you want to ask with a computer. In fact, this is a question you can *empirically answer* with a computer. So you have to ask: why would you *not* utilize a computer when it’s highly appropriate to do so?

EG: At the very least, a computer can provide another perspective, another kind of insight into certain problems.

JS: Parag, you just mentioned your research on North Indian music; can you give an overview of that work and how it relates to some of the issues we’ve already discussed?

PC: I started that work a long time ago as a graduate student at Stanford. I play the *sarod*, (which is a fretless, plucked string instrument), and through learning the sarod, I became interested in a very particular question: ‘What is it that defines a *raag*?’ There are various perspectives on this; some people describe a *raag* in terms of scales or notes. But we actually learn a *raag* in terms of short phrases and techniques for elaborating those phrases. And there’s a complex aesthetic theory surrounding it as well; a *raag* is supposed to connote a mood or a set of moods. But I was interested in using MIR techniques to help figure out the simplest way to describe how we distinguish one *raag* from another. So this may or may not be the way that a human would do it, but what we eventually found was something rather surprising. We learned that by just counting notes, a computer can do a very good job of distinguishing *raags*, even *raags* that have the same scales. And the reason for that, of course, is that if you’re playing certain phrases, you’re going to tend to emphasise certain

notes – you end on some, and you agogically stress others, and so on and so forth. So that was kind of a surprising conclusion, and when I first told my Indian musician colleagues and friends ‘Hey, you can identify raags just by counting notes’, they were definitely surprised.⁶⁷

JS: Emilia, how do these types of ideas relate to your work on flamenco?

EG: Flamenco is an oral tradition and, generally, there are no written scores; in this regard, much of the analytical work on flamenco makes use of recordings. One of the main discussions in this research is about the origin and evolution of flamenco’s different styles. For instance, we ask questions about how certain styles were generated, about whether one particular performance was pivotal in the creation of a new style.⁶⁸ To address these questions, we have been using computational tools to automatically extract melodies, at different representation levels. For instance, one transcription might contain all of a melody’s ornamentations, another only the basic melodic contour. And then we use this representation to compare two performances, establishing connections and distinctions and so forth. When you expand such an approach to a large collection of music, you can really quantify stylistic features. Much of our work has concerned the singing voice. In fact, the voice is perhaps the most challenging instrument to transcribe and describe, because it’s very varied in flamenco; vocal lines contain a lot of ornamentation, and are very rich in terms of expression. So in our experience, it’s another way to attack the problem. For example, we have experts who say ‘OK, this performance is characterized by this type of ornamentation, or by this scale’, and then we try to formalize these perspectives to help automatically extract a representation from a recording. And then we use that representation to compare and to measure the similarity between different performances as a way to understand the problem of flamenco’s ‘styles’. For that, I think, computer tools provide a very good way to formalize expert knowledge. The tools can also be used to assist in the transcription of the music, and to quantify some of the things that are not very easily quantifiable (for instance, vibrato or ornamentation, or small interpretative-deviations that you can analyse very carefully with a computer). [See Figure 18.]

PC: I want to pick up on something important. Obviously, we all know that standard Western notation is a kind of technology, a very brilliant technology that has allowed us to abstract music and disseminate it. Clearly, it’s very good for certain things: it’s good at representing pitch, it’s good at representing durations, things like that. But it’s very poor at representing continuous information, be it subtle pitch deviations, continuous pitch motions, expressive timing deviations, timbre, and so on. So I think part of what’s so powerful about computational tools – and this is exactly what Emilia was saying – is that they allow us to bring to the fore these really important dimensions of music. It is precisely because we’re bad at

67 Parag Chordia and Sertan Sentürk, ‘Joint Recognition of Raag and Tonic in North Indian Music’, *Computer Music Journal* 37/3 (2013), 82–98.

68 Emilia Gómez and Jordi Bonada, ‘Towards Computer-Assisted Flamenco Transcription: An Experimental Comparison of Automatic Transcription Algorithms as Applied to A Cappella Singing’, *Computer Music Journal* 37/2 (2013), 73–90.

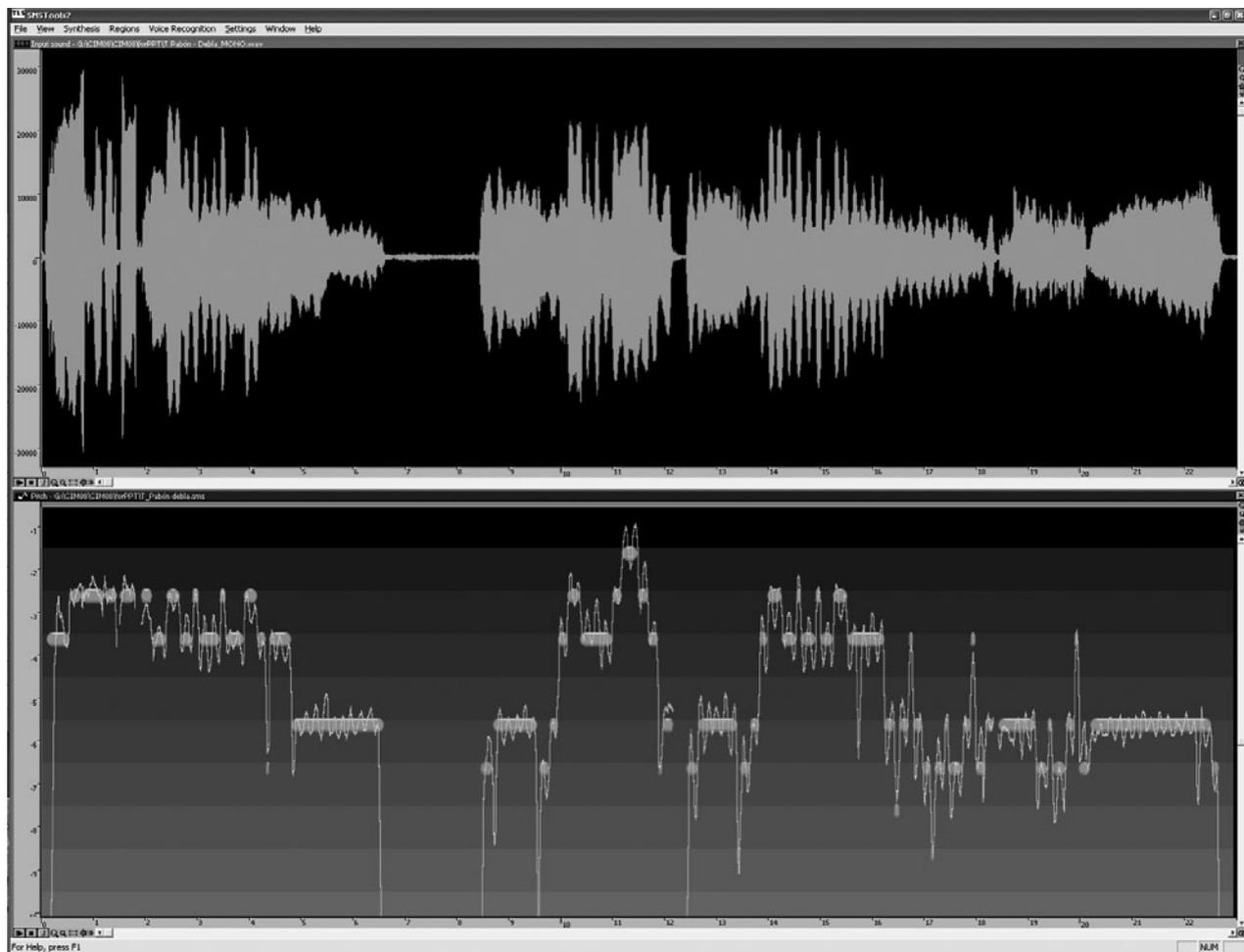


Figure 18 Automatic transcription of the melodic exposition of a *debla* (flamenco cappella singing style) by singer Tomás Pabón. Top: Audio waveform. Bottom: Estimated fundamental frequency envelope and transcribed notes (ovals). Each horizontal line represents a semitone interval. Transcription by Emilia Gómez. Unpublished. Used with permission of author.

certain things that we bring the things we're good at front and centre. So it's not that computational tools right now are perfect, and that they're much better at everything a human could do – not even close; it's almost precisely because they're stupid, that they help us emphasise the crucial things such as timbre that are like elephants in the room.

EG: Computers or communicational models can assist with the transcription problem. But of course, human transcription is not expendable. We should find a way for these two forms of transcription to complement each other. Timbre and microrhythmic deviation are difficult to measure using just the human ear. And there are things that would take ages for a computer to extract from an audio recording that a person might be able to do very easily. That's where the big potential lies: in combining computational models with traditional methodologies, in foregrounding different perspectives on the same problem.

JS: What are some of the challenges in terms of presenting MIR results? I'm thinking specifically about visual representations of the research you do. Parag, I was looking at your dissertation, for example, and I think when people think about transcription – at least in the academic realm – they think about standard Western notation. But you're doing these automatic transcriptions into *bols*. [See Figure 19.] And Emilia, in your recent flamenco research, you have these composite types of visual representations of the outputs from the computer datasets. What are some of the critical issues surrounding visual representation for you both?

EG: The MIR community has put a lot of effort into retrieving and describing music automatically, but we have put less effort into trying to present it visually. There are some proposals on trying to provide or present the descriptions in a visual way: for instance, in our research on flamenco music, we're trying to combine Western notation with neumatic notation.

PC: One of the fundamental techniques that we use is machine learning. And that's just a way of saying that models can learn from datasets. The reality is that whenever you're doing machine learning, visualization is an essential part of building and debugging your system. So that's a little bit different from the question you asked, which is 'when you get your results, how do you present them, what kind of visualizations do you use?' There are actually a *lot* of visualization methods that we regularly use in just doing our basic machine learning work, and I don't think that's necessarily all that different compared to other fields. On the question of how I represent music in my own work, I would say a couple of things. One is that in some cases I don't try to represent it visually. A lot of our work is through synthesis. So let's suppose we're trying to model a vocal style, and we have some hypotheses about how people are connecting the notes. Sure, you can make a visualization where you're looking at a transcription and you're doing a synthetic version based on your model. But it might not give the right perceptual weight to different things. So then you synthesize it and say: 'Okay, now that it's synthesized, does it sound like what I expected to hear? Does it sound different?' So *auralization* is part of the presentation and verification of the results as much as visualization is. For me, the idea of analysis by synthesis is quite relevant.

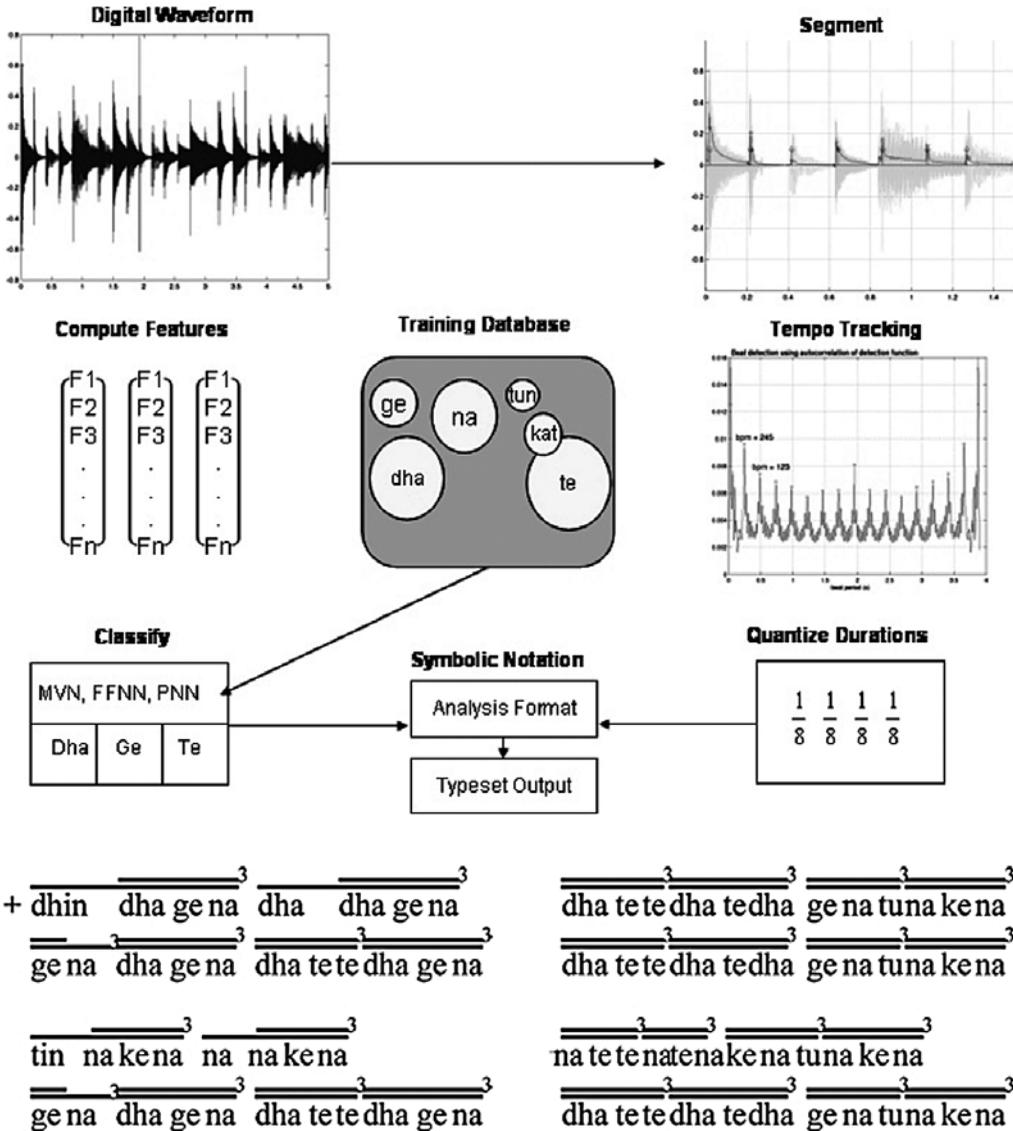


Figure 19 The upper part of the figure shows the main processing steps in the automatic tabla transcription system. First, the audio signal is segmented automatically into individual strokes. For each stroke, audio features are computed which characterize the timbre of the stroke. However, for more complex decisions, a statistical model is used that has been computed from a set of labelled training examples. For each stroke type, the typical timbral values and ranges are computed giving a probability distribution. This can then be used to assess the probability of each stroke type, given an unlabelled stroke whose timbre values have been computed. In parallel, the onset information is used to estimate the relative durations of each stroke. The stroke duration information and the stroke labels are combined to create the symbolic notation. The lower part of the figure shows the output of the transcription system. The symbolic notation is a slight modification of traditional staff tabla notation in which the durations are more explicitly notated, in a manner analogous to Western staff notation. Each continuous group of strokes represents a single beat in the rhythmic cycle. Originally published in Parag Chordia, 'Automatic Transcription of Tabla Music', PhD diss. (Stanford University, 2005), 134. Used with permission of the author.

JS: What about the music scholar who simply wants to do a standard transcription of a song or a piece? More than anything, they just want to get the notes and rhythms on the page. Do you think that this type of person could benefit from looking into MIR research of the kinds you both do?

PC: The short answer is: it depends on how much music they're transcribing. If they're transcribing a small quantity of music, it's going to be faster and easier for them to do it manually, because of the amount of time you'll spend correcting the errors in the transcriptions. Having said that, the calculation starts to change when you're trying to analyse more data. You need to ask: What error rate is acceptable? How many wrong notes and insertions are acceptable? In some cases, a 10% error rate might be fine. In other cases, you really need it to be perfect – at least to your ear. So if you're looking for a more-or-less perfect transcription of a small amount of data, I would say that the current tools aren't that useful, especially if we're talking about polyphonic music.

JS: One of the things that comes up over and over again for music transcribers working in ethnomusicology, musicology, and music theory, is how to render into a visual form really dense amounts of audio information. Parag just brought up the polyphonic issue. Emilia, do you think MIR can ever solve that problem – the disarticulation of dense music textures?

EG: Yes, I think so. There's been progress in this domain. For instance, the state-of-the-art automatic music transcription of maybe ten years ago only worked well with monophonic music signals. At the moment, we have systems that can work fairly well with piano music and guitar music. We are now seeing, for example, systems that can extract the predominant melody, and then the chord progression, and then the instruments that are playing. I think in the near future we will see systems working with big orchestras, extracting the principal melodic lines. All this is due to progress on the audio processing side. For instance, we now have systems for source separation that can attempt to separate the different instruments of a mix.

PC: On that question of 'When is human-level transcription going to be available from a computer with the press of a button?': I think it's coming. The main impediment right now is datasets; it's very time-consuming to label data. Obviously if you already have the score and you're just trying to examine whether a certain performance aligns with the score, then we can do that. But if you're talking about an improvised piece and you have no score, then it's a more difficult problem. We're not there yet, but I think you will be able to get human-accurate transcriptions within five to ten years.

EG: Parag made the very good point that we build our systems by learning from human annotations or from human transcription. Unfortunately, there is not really a readily available corpus of transcriptions done by musicologists or ethnomusicologists. If you want the computer to help you to do the work, you have to first spend some time *teaching* the computer how to do it properly. Because we cannot guess what a feasible transcription would be otherwise.

PC: Especially in oral traditions, where you have a lot of continuous pitch movements, transcription is a pretty fraught question. What does it mean to transcribe such music? One of the things that you're doing when you're transcribing is *interpreting*, right? 'This is important to notate, this is not important.' So, clearly, this question will always need to be there: 'What is my *end goal*?' At the end of the day, there is no magic to transcription. In some cases it's pretty simple; you just want to get the notes and durations. But in a lot of cases it's not that simple. Frankly speaking, in Western popular music, transcribing the melody – what does it even mean? It's very reductive; we're not transcribing all the swoops and dips and so on. In the end, you have to understand why you're doing it. It's not a simple question of pushing a button.

Bibliography

- Abraham, Otto and Erich M. von Hornbostel. 'Suggested Methods for the Transcription of Exotic Music'. *Ethnomusicology* 38/3 (1994 [1909–10]), 425–56.
- Agawu, Kofi. *Representing African Music: Postcolonial Notes, Queries, Positions*. New York: Routledge, 2003.
- Benadon, Fernando. 'Slicing the Beat: Jazz Eighth-Notes as Expressive Microrhythm'. *Ethnomusicology* 50/1 (2006), 73–98.
- . 'Time Warps in Early Jazz'. *Music Theory Spectrum* 31/1 (2009), 1–25.
- . 'Gridless Beats'. *Perspectives of New Music* 47/1 (2009), 135–64.
- Berliner, Paul. *The Soul of Mbira: Music and Traditions of the Shona People of Zimbabwe*. Chicago: University of Chicago Press, 1993 [1978].
- Brackett, David. *Interpreting Popular Music*. Cambridge: Cambridge University Press, 1995.
- Browner, Tara. *Heartbeat of the People: Music and Dance of the Northern Pow-Wow*. Urbana: University of Illinois Press, 2002.
- . *Songs from 'A New Circle of Voices': The Sixteenth Annual Pow-Wow at UCLA*. Middleton, Wis: A-R Editions, 2009.
- Chordia, Parag. 'Automatic Transcription of Tabla Music'. PhD diss., Stanford University, 2005.
- Chordia, Parag and Sertan Sentürk. 'Joint Recognition of Raag and Tonic in North Indian Music'. *Computer Music Journal* 37/3 (2013), 82–98.
- Danielsen, Anne. *Presence and Pleasure: The Funk Grooves of James Brown and Parliament*. Middletown, Conn: Wesleyan University Press, 2006.
- . 'Here, There and Everywhere: Three Accounts of Pulse in D'Angelo's "Left and Right"', in *Musical Rhythm in the Age of Digital Reproduction*, ed. Anne Danielsen. Surrey and Burlington: Ashgate Publishing Group, 2010. 19–35.
- . 'The Sound of Crossover: Micro-rhythm and Sonic Pleasure in Michael Jackson's "Don't Stop 'Til You Get Enough"'. *Popular Music and Society* 35/2 (2012), 151–68.
- Ellingson, Ter. 'Transcription', in *Ethnomusicology: An Introduction*, ed. Helen Myers. New York: W.W. Norton & Company. 110–52.
- England, Nicholas M., et al. 'Symposium on Transcription and Analysis: A Hukwe Song with Musical Bow'. *Ethnomusicology* 8/3 (1964), 223–77.
- Garfias, Robert. 'Transcription I'. *Ethnomusicology* 8/3 (1964), 233–40.
- Gómez, Emilia and Jordi Bonada. 'Towards Computer-Assisted Flamenco Transcription: An Experimental Comparison of Automatic Transcription Algorithms as Applied to A Cappella Singing'. *Computer Music Journal* 37/2 (2013), 73–90.
- Gómez, Emilia, Perfecto Herrera, and Francisco Gómez-Martin. 'Computational Ethnomusicology: Perspectives and Challenges'. *Journal of New Music Research* 42/2 (2013), 111–12.
- Gopinath, Sumanth. 'Reich in Blackface: *Oh Dem Watermelons* and Radical Minstrelsy in the 1960s'. *Journal of the Society for American Music* 5/2 (2011), 139–93.
- . *The Ringtone Dialectic: Economy and Cultural Form*. Cambridge, MA: MIT Press, 2013.
- Griffiths, Dai. 'From Lyric to Anti-Lyric: Analyzing the Words in Pop Song', in *Analyzing Popular Music*, ed. Allan F. Moore. Cambridge: Cambridge University Press, 2003. 39–59.
- . 'Internal Rhyme in "The Boy with a Moon and Star on His Head", Cat Stevens, 1972'. *Popular Music* 31/3 (2012), 383–400.

- . 'Words to Songs and the Internet: A Comparative Study of Transcriptions of Words to the Song "Midnight Train to Georgia", Recorded by Gladys Knight and the Pips in 1973'. *Popular Music and Society* 36/2 (2013), 234–73.
- Herndon, Marcia. 'Analysis: The Herding of Sacred Cows?' *Ethnomusicology* 18/2 (1974), 219–62.
- Hood, Mantle. 'Transcription and Notation', in *The Ethnomusicologist*. New York: McGraw-Hill, 1971. 50–122.
- Jairazbhoy, Nazir. 'The "Objective" and Subjective View in Music Transcription'. *Ethnomusicology* 21/2 (1977), 263–73.
- Jones, A.M. 'On Transcribing African Music'. *African Music* 2/1 (1958), 11–14.
- Klapuri, Anssi and Manuel Davy, eds. *Signal Processing Methods for Music Transcription*. New York: Springer, 2006.
- Kolinski, Mieczyslaw. 'Transcription II'. *Ethnomusicology* 8/3 (1964), 241–51.
- Kroodtsma, Donald E. *The Singing Life of Birds: The Art and Science of Listening to Birdsong*. Boston: Houghton Mifflin, 2005.
- Lanier, Jaron. *You Are Not a Gadget: A Manifesto*. London: Penguin Books, 2010.
- Law, John. *After Method: Mess in Social Science Research*. London: Routledge, 2004.
- List, George. 'Transcription III'. *Ethnomusicology* 8/3 (1964), 252–65.
- Marian-Balasa, Marin. 'Who Actually Needs Transcription? Notes on the Modern Rise of a Method and the Post-modern Fall of an Ideology'. *The World of Music* 47/2 (2005), 5–29.
- Miller, Kiri. *Traveling Home: Sacred Harp Singing and American Pluralism*. Urbana: University of Illinois Press, 2008.
- . 'Schizophonic Performance: *Guitar Hero*, *Rock Band*, and Virtual Virtuosity'. *Journal of the Society for American Music* 3/4 (2009), 395–429.
- . *Playing Along: Digital Games, YouTube, and Virtual Performance*. New York: Oxford University Press, 2012.
- Mundy, Rachel. 'Birdsong and the Image of Evolution', *Society and Animals* 17/3 (2009), 206–23.
- Nettl, Bruno. 'I Can't Say a Thing until I've Seen the Score: Transcription', in *The Study of Ethnomusicology: Thirty-One Issues and Concepts*. Urbana: University of Illinois Press, 2005. 74–91.
- Payne, Roger S. and Scott McVay. 'Songs of the Humpback Whales'. *Science* 173 (August 1971), 587–97.
- Rás, Zbigniew W, and Alija A Wiczorkowska, eds. *Advances in Music Information Retrieval*. Berlin: Springer, 2010.
- Rhodes, Willard. 'Transcription IV'. *Ethnomusicology* 8/3 (1964), 265–72.
- Rice, Timothy. 'Ethnomusicological Theory'. *Yearbook for Traditional Music* 42 (2010), 100–34.
- . *Ethnomusicology: A Very Short Introduction*. Oxford: Oxford University Press, 2014.
- Roth-Gordon, Jennifer. 'Linguistic Techniques of the Self: The Intertextual Language of Racial Empowerment in Politically Conscious Brazilian Hip Hop'. *Language & Communication* 32/1 (2012), 36–47.
- . 'Racial Malleability and the Sensory Regime of Politically Conscious Brazilian Hip Hop'. *The Journal of Latin American and Caribbean Anthropology* 18/2 (2013), 294–313.
- Rothenberg, David. *Why Birds Sing: A Journey Through the Mystery of Bird Song*. New York: Basic Books, 2005.
- . *Thousand Mile Song: Whale Music in a Sea of Sound*. New York: Basic Books, 2008.
- Rothenberg, David, et al. 'Investigation of Musicality in Birdsong'. *Hearing Research* 308 (2014), 71–83.
- Seeger, Charles. 'Prescriptive and Descriptive Music Writing'. *Musical Quarterly* 44/2 (1958), 184–95.
- . 'Report of the Chairman-Moderator'. *Ethnomusicology* 8/3 (1964), 272–77.
- Shelemay, Kay Kaufman. 'Notation and Oral Tradition in Africa', in *The Garland Handbook of African Music*, ed. Ruth Stone. New York: Garland Publishing, 1999. 24–43.
- Solis, Gabriel. 'Thoughts on an Interdiscipline: Music Theory, Analysis, and Social Theory in Ethnomusicology'. *Ethnomusicology* 56/3 (2012), 530–54.
- Stanyek, Jason and Fabio Oliveira. 'Nuances of Continual Variation in the Brazilian Pagode Song "Sorriso Aberto"', in *Analytical and Cross-Cultural Studies in World Music*, ed. Michael Tenzer and John Roeder. New York: Oxford University Press, 2011. 98–146.
- Tenzer, Michael. *Gamelan Gong Kebyar: The Art of Twentieth-Century Balinese Music*. Chicago: University of Chicago Press, 2000.
- . 'Integrating Music: Personal and Global Transformations', in *Analytical and Cross-Cultural Studies in World Music*, ed. Michael Tenzer and John Roeder. New York: Oxford University Press, 2011. 357–87.
- Titon, Jeff. 'Knowing Fieldwork', in *Shadows in the Field: New Perspectives for Fieldwork in Ethnomusicology*, ed. Gregory Barz and Timothy J. Cooley. New York: Oxford University Press, 2008. 25–41.
- Winkler, Peter. 'Writing Ghost Notes: The Poetics and Politics of Transcription', in *Keeping Score: Music, Disciplinary, Culture*, ed. David Schwartz, Anahid Kassabian, and Lawrence Siegel. Charlottesville: University of Virginia Press, 1997. 169–203.