

## CHAPTER 3 – THE SPEECH TRANSCRIPTION PROCESS

Speech is a monophonic and a linear practice<sup>1</sup>. There are generally no pitch relationships from one note to the next. Frequencies are determined by the words spoken and the intention behind what the speaker is saying. A transcriber must first assign the most relevant pitch according to the western tuning system<sup>2</sup> and then form a conclusion as to what the appropriate points of commencement and departure<sup>3</sup> in each sentence are. An experimental process will take place and these sentences will have a pitch<sup>4</sup> and a grouping scheme assigned to them (see Chapter 5) adding to their perceived musicality and instrumentalism.

### Part 1: The Selection Process

#### i) *Why transcribe a speech?*

Moran points out how a Bud Powell line requires a different technique to that of a laugh<sup>5</sup>. Moreover, 'by setting a transcription task you want to be forced to play something or find ideas that you're not used to playing<sup>6</sup>'. Thus, 'by transcribing music, regardless of the style, one's ability to improvise will improve greatly<sup>7</sup>. Transcribing speech would mainly be used to 'escape the gravity of musical style'. O'Connor<sup>8</sup> mentions that the improviser can 'develop new approaches to phrasing' by exploring the human voice.

---

<sup>1</sup> Monophonic: a single musical line. Sometimes two monophonic melodies will be spoken simultaneously.

<sup>2</sup> Equal Temperament, the 12-tone chromatic scale, Edward M Burns 1999, *Fyk* "(1982a, 1982b)

<sup>3</sup> Accented notes or principal tones

<sup>4</sup> In relation to the problem of speech not containing notes from Western pitch, a comparison could be made to *ragas* found in Indian classical music whereas the *shrutis* or microtones are slight variations of certain intervals, the exact values of which are dependent on the individual melodic framework *raga* being played (Burns, 1999). Each speech therefore must be analysed/transcribed based on the system of notes, and not in isolated cases

<sup>5</sup> Moran, 2018

<sup>6</sup> *Ibid*

<sup>7</sup> Hannaford, 2017

<sup>8</sup> O'Connor, 2018

Neville explores other features of the voice such as the ‘mood and emotion of a speech excerpt, as opposed to its rhythmic and /or melodic qualities’<sup>1</sup>. Moran goes further explaining the curiosity of hearing sounds and tones.

“ What is the tone of the phrase? You know, a loving tone vs. an angry tone”<sup>2</sup>

O’Connor<sup>3</sup> describes a good reason for transcribing speech, ‘there is a considered approach to contour and inflection that is independent of musical syntax and convention’. This transcription would mainly be used to ‘escape the gravity of musical style’.

ii) *Good Speech, Bad Speech*

Moran says he does not want to transcribe anything that he does not want his kids to hear<sup>4</sup>. He does not want to ‘cross the line to certain kinds of speeches because that’s not what I’m here to promote’<sup>5</sup>. In the transcriptions of Donald Trump, the Adolf Hitler excerpt, and the Pauline Hanson speech/podcast, while the language isn’t promoted, I am touching on the parts of speech that are, to use Moran’s term ‘a little bit gross’. It is interesting musically to focus on the negatives as well as the positives as music needs to contain consonance and dissonance for it to have a healthy balance.

iii) *Unlocking speech*

Moran briefly explains that there are ‘language codes’ in music as well as in speech, such as in Schoenberg’s ‘Pierre Lunaire’ or Wagner’s ‘Ring Cycle’. The code is aimed at an audience who have the potential to unlock it, and not aimed at people who may disagree with that code. Trump’s ‘Build a Wall’<sup>6</sup>, is perhaps code for something far more relatable to his followers. In other words, codes and metaphor can have the

---

<sup>1</sup> Neville, 2014

<sup>2</sup> Moran, 2018

<sup>3</sup> O’Connor, 2018

<sup>4</sup> Moran, 2018

<sup>5</sup> Ibid

<sup>6</sup> I.e. A Donald Trump speech, Moran does not want to even know what it feels like to play the speech ‘in his hand’. Moran, 2018

potential to be more powerful than messages that are literal because ‘they can be more visceral, and more emotional’<sup>1</sup>.

## Part 2: Methods of Transcribing

### i) Introduction

In order to learn how to play a piece or how to copy elements from a section of a piece, an artist may listen to versions of the same piece played by another artist that they want to play themselves. For instance, if one wants to learn the tune *Autumn Leaves*<sup>2</sup>, the Keith Jarrett Recording live at the Blue Note<sup>3</sup> would be a great reference recording for the following reasons. First, Jarrett takes a 32 Bar A-A-B form and develops many structural and melodic permutations with added ostinatos that although staying in the temperament of the piece, transforms it into a piece of epic size (26 min). In Jarrett’s opening introduction he plays in many different tonal centres and uses fragments of the melody to navigate through the improvisation. This enables the listener or the transcribing artist to discover seemingly endless possibilities that an artist can utilize<sup>4</sup>.

**Autumn Leaves Solo**  
Keith Jarrett/Gary Peacock/Jack DeJohnette  
Trans. Matt Robbins

The musical score for 'Autumn Leaves Solo' is presented in a three-staff format. The top staff is for Piano, the middle for Acoustic Bass, and the bottom for Acoustic Bass. The tempo is marked as quarter note = 100, with a swing feel and a duration of 4:22. The key signature is two flats (B-flat and E-flat). The piano part includes chords: Cm9, F13, Bbm7, E9, and Eb9. The acoustic bass part starts with a bass line marked with an '8' below the first note. The melody is marked with 'tritone sub 2-5' above the final measure.

Figure 3-1 Autumn Leaves melody

<sup>1</sup>Ibid

<sup>2</sup> Joseph Kosma, originally 'Les Feuilles mortes'

<sup>3</sup> Jarrett, Keith. *Keith Jarrett at the Blue Note*. ECM, 1995. Recorded in 1994.

<sup>4</sup> Of the trio's other "Autumn Leaves" recordings, only 'Up For It' (2002) features a similarly expansive treatment.

Jarrett maintains a communicative element to his playing throughout the piece. He achieves this by playing simply and melodically (Figure 3-1). When one transcribes Jarrett one quickly discovers that something that sounds easy, is deeply complex, and difficult to transcribe, memorise and play. This is one of the reasons aspiring improvisers should engage in a process that episodically and cyclically turns the complex into the simple and vice versa.

Transcribing<sup>1</sup> ultimately enhances the professional or aspiring artists inner pitch, articulation and rhythm comprehension and through copying and emulating the masters of musical communication<sup>2</sup> the artist can thus communicate his or her own ideas to their audience in a more deliberate fashion. So, what about speech?

Transcribing the spoken voice is simply another way to further ones understanding of their own performing limits, especially as an improviser<sup>3</sup>. This means that to transcribe a speech one needs to focus on the overall context of the tuning or placement of the notes to properly fathom what would be the best way to play it on an instrument.

Moran looks at the transcribing process through layers/steps providing a strong method for his speech studies to grow organically into compositions of their own. So, how can one transcribe more efficiently?<sup>4</sup> That depends on what the transcriber wants to achieve<sup>5</sup>.

ii) *The segmental listening/writing method*<sup>6</sup>

This method uses the program *Amazing Slow Downer* or *Transcribe* for easy looping of the segment. These programs enable the transcriber to adjust speed without

---

<sup>1</sup> Relative pitch possessors can identify notes that are “out of tune” intervals. Burns and Ward, 1978

<sup>2</sup> ‘Communication comes from a place where both parties share a common understanding of what is meaningful, i.e. language, past experiences, previous interactions with others’, Hannaford, 2018

<sup>3</sup> ‘By setting yourself this task you want to be forced to play something or find ideas that you’re not used to playing’. Burns, p 234, ‘It is clear, that the context had a significant effect on both the “discriminability” and by the subjective sizes of the intervals’

<sup>4</sup> Hannaford talks about the time taken to address the problem of transcribing and concludes that it does not matter how long it takes to get a result, ‘It’s just about the outcome’. Hannaford, 2018

<sup>5</sup> ‘The improviser is aware on the first hearing, of the exact pitches used by other performers’. Coker, 1964

<sup>6</sup> See segmental listening/writing method video

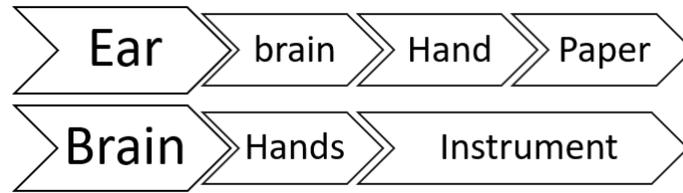
compromising the pitch quality. You will need a score and a pencil to write (scribe) the notes and rhythm down. *Sibelius* or *Finale* are great tools in writing notes on a PC or MAC. Do this exercise without an instrument and use the following steps:

- A. Choose 2 or 3 note segment/fragments to loop at 25% of the original speed and start writing the notes down. The rhythm will come later.
- B. Continue this process until all the pitches of the segment that you wish to transcribe have been written.
- C. Once the notes/pitches have been written, expand the listening scope to around 8 notes.
- D. Listen to these 8 notes at 50% speed or whatever speed is most beneficial to hear the rhythms correctly.
- E. Think of a metric sub-division other than the main pulse of the piece to truly and mathematically make the rhythm correct. For example, if the piece deals mainly in swung quavers (8<sup>th</sup> notes) think in semiquavers (16<sup>ths</sup>). If the rhythm does not fit into 16<sup>th</sup> divisions try 32<sup>nd</sup> divisions. This will help to negotiate and comprehend any irregularities in the rhythm. If this does not work, listen for triplets in any numerical value<sup>1</sup>.
- F. Once the rhythm and pitch are correctly written the instrumentalist can play along with the precise score.

This method is effective from the first listen to the time taken to play the fragment perfectly, especially for a good reader and good listener. The negative to this method is the speech might take longer to internalise, and the transcriber might not develop his/her ear training and communication skills with an audience when improvising. This is because the method doesn't use his or her instrument in the process (See Figure 3-2)

---

<sup>1</sup> 'Jason Moran meets Hue Blanes' is a transcription with the most accurate rhythm using conventional notation to date.



**Figure 3-2** Segmented listening/writing method (top) vs. repetition ear method (bottom)-Coker

iii) *The repetition ear method*<sup>1</sup>

This method is all about listening to build your ear. Moran says,

“ ‘So, when I wasn't having it in my hand at the piano that I at least had it in my ear, and I could then start to find a nuance’<sup>2</sup>. ”

This method uses the program *Amazing Slow Downer* or *Transcribe* for easy looping of the segment. You will need your instrument or voice to complete this exercise.

- A. Choose a segment/fragment to loop and start playing along with it at about 25% speed. It will be difficult at first to be exact and may take days or weeks to get the notes precise.
- B. Slowing down the audio (25%) for transcription does sometimes present its own problems. ‘Slowing down the audio revealed rises and falls in inflection that led to two or more notes being present in a single syllable’<sup>3</sup>.
- C. Gradually add speed to the fragment. You will notice once again that the fragment will be difficult to play.
- D. Continue increasing speed until you get to the original tempo<sup>4</sup>.
- E. Do not write out the transcription at any stage of this process.

Once this method is attempted the transcriber can then move onto the next fragment when a desired section of the piece is internalized. ‘I was doing everything by ear and memorizing the piece straight away (*repetition ear method*) whereas previously in my undergraduate degree I would transcribe with paper and a pencil

<sup>1</sup> See video, Repetition Ear Method

<sup>2</sup> Moran, 2018

<sup>3</sup> Neville, 2014

<sup>4</sup> “Many an hour was spent ‘note bashing’ on a piano just to achieve approximations of some notes and phrases, people don’t talk in an equally tempered tuning system” Neville, 2014

and I would write it down like a piece of classical music (*segmented writing method*) and then try and try and then try and play along to the recording'<sup>1</sup>.

Internalisation means that the number of repetitive listens to a recorded fragment develops that music language inside of the transcribers mind/ear and becomes a natural part of their musicality in other situations including improvisation. The *repetition ear method* can take a long time to hear the pitches and rhythms of a Jarrett (*Repetition ear method*) or especially a Dolphy (*segmented listening method*) and can be frustrating when the playback speed is too rapid for any level of real-time comprehension.

iv) *The reading method*<sup>2</sup>

This is the fastest method between hearing a recorded solo and being able to play it on an instrument. It works as follows,

- A. Acquire a score of a previously completed transcription.
- B. Play the score without the recording at your own desired tempo, with the idea that you can build tempo gradually over time.
- C. Try to play along with the recording.

The *reading method*, with conventional notation, enables the transcriber to visually see the shapes of the rhythms and pitch contours used by the improvising performer. If the 'transcriber' is a good note reader, he/she can play it near perfect at first glance. The downside to this method is the possibility that no internalization<sup>3</sup> is taking place in the performers' development as an improvising artist. This is unless he/she memorises the piece, and or chord structure.

v) *The singing method*<sup>4</sup>

This method only requires *Transcribe* or *Amazing Slow Downer*, and a voice. It consists of the following steps,

---

<sup>1</sup> Hannoford, 2017

<sup>2</sup> See video, The Reading Method

<sup>3</sup> *Internalising* and *Memorising* are not the same ideas. Memorising enables a player to play a piece without sheet music. However, internalising the music requires a broader knowledge of the piece i.e. harmony, melody, structure, in order to improvise if necessary.

<sup>4</sup> See video, The Singing Method

- A. Listen a few times to a piece/speech that you would like to transcribe.
- B. Sing along with the phrases as accurately as possible, repeat<sup>1</sup>.
- C. After being comfortable with the phrases in the piece, sing it without the recording

This exercise is simple but effective. An internal knowledge of the phrase is evident but transferring this knowledge to your instrument is where the difficulty will arise.

vi) *The real time self-transcription method*<sup>2</sup>

This method consists of a sentence of a self-transcribed speech harmonised and restructured to mimic the voice, done in real time. 'There's something about understanding where your ear is and where in relationship to the melody it is'<sup>3</sup>. This method goes as follows,

- A. Record a sentence, a shorter sentence will ensure quicker progress
- B. Play along to a loop of the recording with no real aim
- C. Find the melody of the sentence
- D. Create a harmonisation of the melody
- E. Find the 'flow' of the voice<sup>4</sup>
- F. Play along with the sentence, put chords to it, mimic it, make fun of the imperfections of your own voice<sup>5</sup>. Vary the playback speeds.

The user can learn about their own voice and about their playing style instantly. The user may find new ways to harmonise melody. This method is the shortest way to transcribe a voice without writing anything down. The transcriber will be able to build their ear training skills quickly and effectively.

The *real time self transcribing method* can be frustrating as it is a spontaneous exercise. A lot of mistakes in melody recognition and harmony implementation can be made. To counter this problem, listen back to the recording in as slow speed as

---

<sup>1</sup> Olivia Chindamo, Secret Love, <https://www.youtube.com/watch?v=tdfTYh4vUg0>

<sup>2</sup> See video, The Real Time Self Transcription Method

<sup>3</sup> Moran could be making a comment about bringing the ear ever so closer to the melody of the speaker, even in real time, thus quickening his/her melodic speech recognition skills

<sup>4</sup> Rhythmic contour

<sup>5</sup> 'Australians have a real distinct way of talking like a lot of different people. So, there's maybe something about the way I speak and all the people around me speak that could give me some kind of Australian you know not Australian thing but relate to being Australian'. Hannoford, 2017

possible. You may not like the sound of your own voice. This needs to be overcome both in your voice and in your playing. An improviser must be positive about what they are playing in the moment.

### Part 5: Summary

The methods explored in this chapter are based on personal experiences and experiences of my interview subjects. These methods are useful in transcribing speech. Ideally each method would be used concurrently. The ‘Things That Have Been Said (TTHBS)’<sup>1</sup> project primarily used the *segmented/listening* and the *writing method*. In retrospect, it would have been better to further my ear training skills by utilising other methods of transcription, particularly the *repetition ear method* and the *real time self-transcription method*. It was not a desire at the time to internalise or memorise the speeches as there were time constraints, although the simple act of playing the pieces over and over whilst reading the score enabled me to internalise the speeches in some capacity (*repetition ear method and reading method*).

An inspiration for the *repetition ear method* and the *real time self-transcription method* came from a quote from Mark Hannoford<sup>2</sup>

“ I would read the most banal shit like washing instructions on a jumper or something, record it, and then try and play it over and I'm pretty sure for those recordings I was using my laptop and I was always trying to do it at real speed and never slowing things down. ”

During our interview, Moran was very clear about his method of transcribing

“ Moran (M): ‘Well I've been using *Logic* (Micro Logic<sup>1</sup>) for a very very long time now and *Logic* has always been my saviour. So, I would put on the

<sup>1</sup> ‘Things That Have Been Said’ TTHBS June 2017 Melbourne International Jazz Festival Premiere

<sup>2</sup> An interesting observation of Hannaford is he chooses not to slow the transcription down but to transcribe at ‘real speed’. This would add stress to the transcriber. However, Hannaford observes that he has formerly used the *segmented listening/writing method* in his ‘undergraduate’ years and has evolved to confidently relying on his ear (*repetition ear method*) to process information.

sound file first, I would record the sound on to Minidisc which has always been my best friend

Blanes (B): Do you still have it (the Minidisc)?

M: 'I still do (laughs). Then I would transfer it onto the computer and then I would put it in *Logic*. In *Logic*, I would just go phrase by phrase and I'll put it on loop 2 seconds of a time then I would play it on the keyboard then I'll move to the next 2 seconds and I'll go through the whole piece and it might take 3 or 4 hours or something. And that was just the first layer. ”

And

“ M: Secondly, *Logic* does a really bad job of transcribing what you play especially back then. So, it was just more kinda looking at it to find the notes again and then trying to play it in time with the voice and trying to get the inflections and that's when the technique started to get challenged. ”

Moran<sup>2</sup> uses the software, *Logic* to split the difference between my two transcribing methods<sup>3</sup>, *the segmented listening/writing method*, and *the repetition ear method*. All methods are a great way to study 'your own voice on the instrument'<sup>4</sup>. Both Moran and Hannofard mention the importance of this in their own findings.

---

<sup>1</sup> Computer Software program that writes down the notes (not always accurate) as you play them on a MIDI keyboard

<sup>2</sup> "Since his emergence on the music scene in the late 1990s, Kennedy Center Artistic Director for Jazz **Jason Moran** has proven more than his brilliance as a performer. With a unique vision and innovative approach to the music, Mr Moran was appointed Artistic Advisor for Jazz for the John F. Kennedy Center for the Performing Arts in November 2011 and given the title of Artistic Director for Jazz in May 2014", taken from <http://www.kennedy-center.org/artist/A4757>

<sup>3</sup> Moran transcribes segmentally and compartmentally, and he plays the phrase physically into his *Logic* keyboard in order to transcribe/write the notes into the computer while playing. Moran, 2018

<sup>4</sup> *Ibid*