SYNTHRUMENTATION, REVISITED:

TOWARDS A NEW METHOD OF ADDITIVELY SYNTHESISING SPEECH

IN ACOUSTIC INSTRUMENTAL CONTEXTS

by

ANDREW CHEN

A dissertation presented to the Research Faculty of the Royal College of Music,

in partial fulfilment of the requirements for the degree of Master of Music in Composition

Royal College of Music

2021

ABSTRACT

Synthrumentation is a term coined by composer Clarence Barlow to refer to his innovative, personal compositional practice of additively (re-)synthesising phonated speech sounds purely via means of acoustic instruments, *i.e.* without the presence of an actual human voice or electronics. I repurpose 'synthrumentation' as a general term (broadly referring to acoustic additive instrumental synthesis of all kinds), and detail a number of extant works utilising it, for the purposes of identifying potential trends and deficiencies within historical practice in terms of intent, pre-compositional procedure, and musical usage. It becomes clear that, amongst other things, the *inharmonicity* that part-constitutes both speech and other sounds has been largely overlooked in existing synthrumental practices, suggesting a fruitful area of new inquiry and potential basis for developing novel synthrumentation methods.

With respect to the above review of synthrumental works 'after the fact,' I continue on to investigate various aspects of sonic material itself, prior to undergoing synthrumental processes, with a particular emphasis on human vocality, phonemes, formants and the characteristics of whispered speech.

Finally, a new method of synthrumentation is devised, aiming to replicate whispered speech, particularly through the continuous formant bandwidths on which it is structured (as opposed to the discrete harmonic peaks of its phonated counterpart), with particular attention given to practical considerations and feasibility in live performance contexts. Resultant orchestrations and audio samples (recorded live by a group of violinists) are appended. Ultimately, the method proves to be successful in fulfilling its base aims, and resultant sounding synthrumentations are, superficially, sonically characteristic and impactful. However, further research is required to objectively determine their potential uses in musical contexts, broader aesthetic implications, and general efficacy.

ACKNOWLEDGMENTS

All being well, this looks to be my final piece of academic work at the College as a student, having been here far, far, far too many years already. Although this is, on the face of it, just a relatively minor MMus-level paper, I'd still like to take the liberty here of indulging in a fairly lengthy, drawn out expression of gratitude to the friends, mentors, and family who have helped me get here so far.

Firstly, thankyou to Ivan Hewett, my research supervisor for this piece of academic work, whose ability to maintain both enthusiasm and an impartial critical eye in our conversations has been an incredible boon to the process, and whose infectious love of music and musicology has been truly inspirational. My thanks also to Julian Anderson, who, unknowingly through our conversations years ago, inspired me to pursue (my then embryonic and generally quite ignorant) concepts of acoustically-synthesised-speech in a far more rigorous and dedicated way, and whose wealth of knowledge and general support of my endeavours has been an incredible help to date. I owe my thanks too to Clarence Barlow, whose work has served in part as an impetus for this project, and for his interest in and support of the research herein.

Thankyou to Sygyt Software, the developers of Voce Vista Video, my spectral analysis program of choice – and one that, although ostensibly designed for singing pedagogues, has been incredibly useful in both my scholarly and compositional work.

My sincere thanks to (from left to right, in positional arrangement) Dimitra, Lara, Viviane, Tony, Ugne and Ella, for playing in the synthrumental formant recording session, and being lovely people and great violinists generally. Thankyou also to the RCM Studios team, particularly Steve Harrington, the recording engineer on the day, and Danny Holland, who helped to facilitate this entire end of the project and has generally put up with my ceaseless, overbearing enthusiasm and presence since my earliest years here as an undergraduate. I also want to acknowledge the countless number of fellow students and friends who have individually donated their time and effort towards helping me realise my compositional ideas over all these past years; in a way, this paper is a culmination of much of my previous work, and they have all played a significant role in the journey along the way.

Thankyou to Alison Kay, Simon Holt, and Jono Cole, my composition teachers over the past six years, who have played a crucial, foundational role in my personal and artistic growth, and who have been a constant source of support and encouragement, through thick and thin. Without them I wouldn’t be the composer I am today, and without them I also wouldn’t know the composer that I want to grow to be, in the days, weeks and years to come.

Finally: thankyou to my family, who have been so incredibly supportive of my musical endeavours since a young age, and to whom I send my love – looking forward to hopefully seeing you all in person again soon. And thankyou (for the second time) to Ella, my partner, proofreader, creative sounding board, and reluctant non-salaried life coach.

Table of Contents

List of Figures

N.B. unless otherwise specified, all spectrograms within this paper have been generated with respect to audio input sourced by the author, using Sygyt Software's Voce Vista Video.

INTRODUCTION

Composers have long been fascinated with the inner workings of sound, and have sought to develop means by which they can control, manipulate, extend, and attenuate it for musical effect. This has prompted some to look beyond conventional conceptions of instrumental function and orchestration in their creative work, or to seek to invoke dislocated and decontextualised timbres entirely, both abstractly, as well as in explicit ways: particularly, through the embedding of discrete sonic objects into a musical fabric.

One such application of this concept can be found in the work of composer Clarence Barlow: in the late 20th century, he devised a new method of "additive synthesis through musical instruments" with the explicit aim of "[approximating the] reproduction of speech sounds solely [acoustically]," *i.e.* without the presence of electronics or an actual human speaker;[1] Barlow later coined the term *synthrumentation* (a portmanteau of 'synthesis' and 'instrumentation') to refer to this personal, spectral technique.[2] In a most basic sense, the synthrumental method, taking the spectrograms of a *target* speech audio material as a reference in the first instance, transcribes and reorchestrates its harmonic components whilst discarding or otherwise leaving aside their inharmonic counterparts;[3] it was applied in its "most fully realised" form in Barlow's chamber orchestral work of 1981-84, *Im Januar Am Nil*.[4]

---

[1] Clarence Barlow, "On the Spectral Analysis of Speech for Subsequent Resynthesis by Acoustic Instruments," *Forum Phoneticum 66: Festschrift Georg Heike* (1998): 184.

[2] Tom Rojo Poller, "Clarence Barlow's Technique of 'Synthrumentation' and Its Use in *Im Januar Am Nil*," *Tempo Vol. 69 No. 271* (2015): 8;

Clarence Barlow, "On Music Derived from Language." International Journal of the International Institute for Advanced Studies in Systems Research and Cybernetics, Vol. 9 No. 1 (2009): 31.

[3] See 1.6. for greater elaboration on harmonicity and inharmonicity in this context.

[4] Poller, "Clarence Barlow's Technique of 'Synthrumentation,'" 9;

Explored further in 1.1 and 1.9; Barlow also made use of this technique in the later *Orchideæ Ordinariæ* (1989), a fact which will be discussed at a later point.

Although Barlow's 'synthrumentation' is procedurally and, on the whole, aesthetically unique, it still shares a number of commonalities with techniques employed by other composers in their work, particularly in that it explicitly invokes discrete, pre-existing sonic objects external to the musical composition in the first instance. This is true not only of *Im Januar*'s most immediate predecessors (or contemporaries) in *spectralism*, but also of works spanning many eras prior to the advent of spectrographic analysis technologies.[5] However, in spite of this apparently substantial foundation, notable musical attempts to non-electronically and purely-instrumentally synthesise human speech subsequent to *Im Januar* are uncommon – both broadly, and within Barlow's oeuvre itself.[6] This absence, as well as the preciseness of procedural focus that Barlow opted to take in *Im Januar* in the first place, suggests that there are unexplored aspects of non-electronic vocal instrumental synthesis still yet unexplored, and potential for discovery of new knowledge and new synthrumentation methods within.

With this in mind, my aims in this investigation are to – under a generalised definition referring to *any* analytically-derived orchestration that aims to instrumentally and non-electronically synthesise pre-existing sonic objects[7] – interrogate what *synthrumentation* constitutes (or might constitute) as a whole beyond Barlow's original singular concept; this will entail a wide-ranging, multidisciplinary approach, viewing synthrumentation not only in terms of compositional procedure, but also historical lineage (including reference to adjacent practices that synthesise non-vocal sonic objects), acoustic, psychoacoustic, and technological considerations, and so forth – so that, ultimately in turn, unexplored areas of

[5] See 1.2.
[6] Works written after *Im Januar* that invoke human vocality by electronic means will also be discussed; see 1.8.
[7] My use of 'orchestration' is interchangeable here with composition and compositional material; this more liberal definition opens up not only the possibility of tracing a more clearly interconnected lineage of similar compositional concepts throughout history, and may also aid the closer identification of ideas, areas, and methods of practice not yet explored within.

I owe all credit for this proposition to Julian Anderson, who suggested to me the idea of generalising *synthrumentation* as such a useful term in the first instance.

practice can be identified, and new synthrumental methods or methodological entry-points can be devised and implemented. Any such methods discovered in this dissertation will be realised in both theoretical, written form, as well as in live-recorded audio examples.

New methods devised will be limited to the use of human vocality as a target for resynthesis from the outset, as a matter of scope. This is on one hand most simply suggested by Barlow's original focus in *Im Januar*; on a more personal note, too, as a composer myself, I am very interested in the deployment of non-electronic synthetic vocality in musical contexts and artistic potential therein, as well as its relation to broader, recent developments in the field of speculative-materialist sonic philosophy (notably regarding concepts of 'entity' and 'agency' within music as a phenomenon distinct from human actors).[8] I hope that this research might be of preliminary value to practice across these fronts and beyond, even if it does not attempt to meaningfully encompass them here.

Do note, too, that this ultimate limitation of scope to human vocality is simply a matter of preference, not absolute necessity – many other examples of mimicry or formalised additive instrumental synthesis of non-vocal target sources do, of course, exist; importantly, although they are treated peripherally here, I would still emphasise their equal value in the event of any broader search for new entry points into synthrumentation beyond just this paper.

[8] Don Ihde, Listening and Voice: Phenomenologies of Sound, 2nd Ed (Albany: SUNY Press, 2007), 185-190;

Holly Watkins, *Musical Vitalities: Ventures in a Biotic Aesthetics of Music* (Chicago: The University of Chicago Press, 2018), 21-23;

See also Christoph Cox's *Sonic Flux*, which although ostensibly a philosophically new work, in part seeks to reify the interlinked importance of earlier musical materialist concepts as exemplified in Cage, Brecht, Lucier, Amacher, Radigue, *etc.*;

Also see *Realism Materialism Art* eds. Cox, Jaskey and Malik, which provides a broad-but-relevant range of perspectives, particularly regarding divergent conceptions of art away from idealism, which have been accumulating at increasing pace over the past two decades.

With regards to methodology – although the nature of this work is exploratory and, in terms of its desired outcome, rather 'experimental,' my aims are explicitly *not* to attempt any rigorous qualitative assessment of its results as above.[9] Such an approach (in requiring, amongst other things, an ethics framework and subjective trials) would stray too far from the remit of this paper, and in so doing detract from its core conceptual focus; some limited reference will still be made, however, to aspects of psychoacoustics where relevant (particularly regarding conceptual bases for devising new methods and questions of 'effectiveness').

Finally, and perhaps most pertinently: the methods and audio examples that result from this research are treated as an ends unto themselves; no doubt they ought to pave the way, and are intended, for use in actual musical compositions, but that is a task best left for future endeavours.

This paper will be divided into three Chapters:

Chapter 1 will, using the seminal *Im Januar* as a focal point, provide an account of Barlow's original synthrumental procedure, and then situate it in the context of a selection of other 'synthrumental' or otherwise conceptually adjacent musical works throughout history. These will be discussed in roughly chronological order, with the direction of discussion motivated by the overarching research aims; a subsequent review of *Im Januar* will, amongst other things, suggest what overlooked aspects or deficiencies of current practice might exist.

Chapter 2 will be dedicated to further investigation of speech, this paper's synthrumental target, in non-musical contexts, integrating a number of sources from across a range of disciplines – including linguistics, acoustics, neurobiology, engineering, and

[9] *i.e.* attempting to determine the extent to which a given synthrumentation technique accurately simulates speech, *etc.*

combinations thereof – with discussions guided by the outcomes of Chapter 1 (and resulting refinement of focus). Particular attention will be paid as well to determining what aspects of speech need to be considered in order for a new synthrumentation to truly be 'effective.'

Finally, in Chapter 3, I will devise a novel method of synthrumentation with reference to all the above, presented both in written explanatory technical detail, as well as in the form of live-recorded audio samples, appended. This will be followed by a brief, informal, subjective discussion.

CHAPTER 1.

SYNTHRUMENTAL LINEAGES

1.1. Barlow's synthrumentation: a brief overview

Rather than arriving as a singular fully-formed concept, Clarence Barlow's synthrumentation was developed over several years of theoretical and practical work from 1981 onwards. In 1997 he authored a paper summarising the procedure and various technical considerations of the method,[10] with particular reference to *Im Januar am Nil*, where its usage was best exemplified.

As laid out in the Introduction, synthrumentation can broadly be thought of as an orchestration technique whereby acoustic instruments 'additively synthesise' a target source – particularly, in Barlow's method, through the transcription and redeployment of pitch material as found in said target source's spectrographic representation (see Fig. 1.1).

[10] Republished in 1998 as cited.

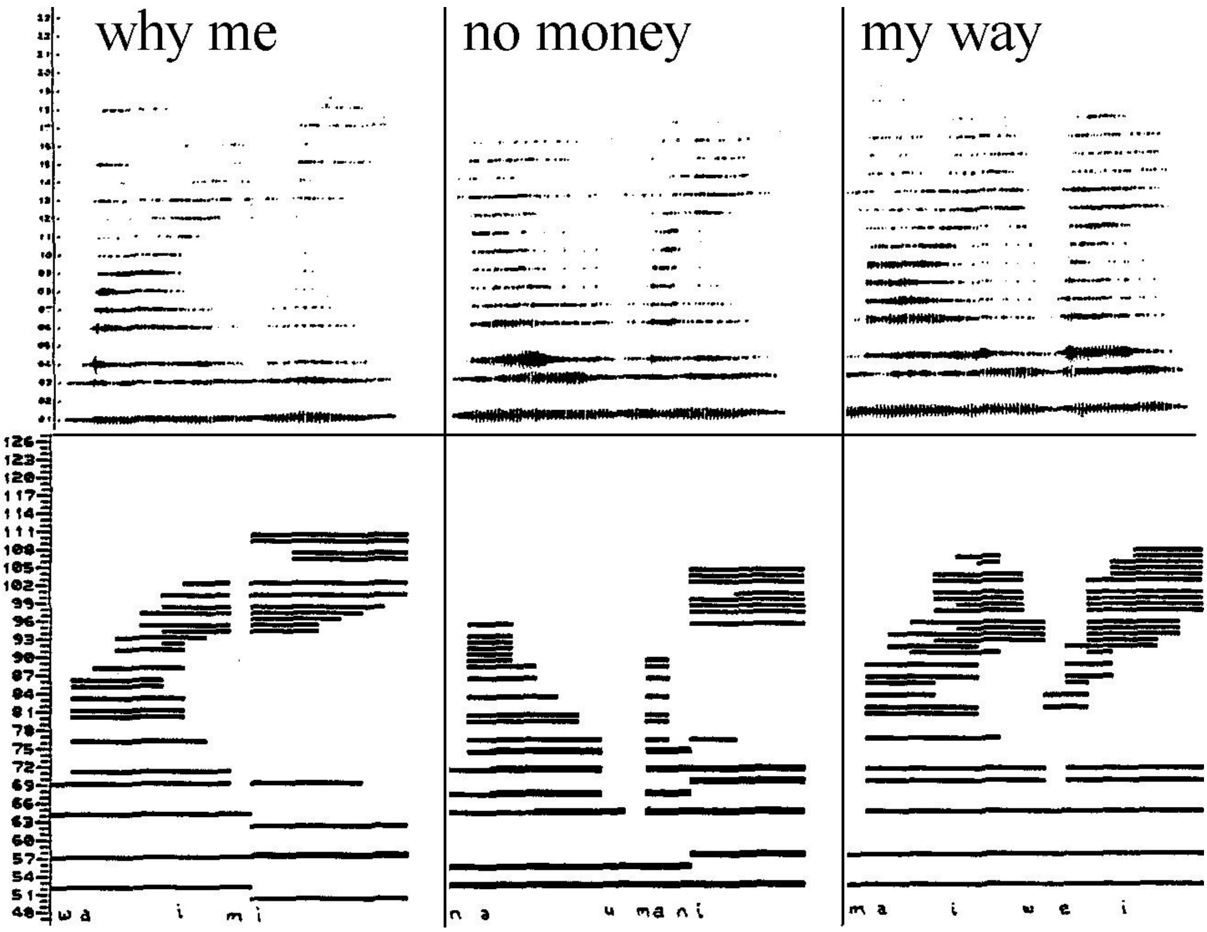


Figure 1.1. Visual demonstration of the spectrogram-transcription method employed in synthrumentation – essentially the realisation of partials as MIDI notes – as reproduced in Barlow's *On Music Derived from Language* (2009). N.B. this transcription actually specifically corresponds to Barlow's later *Orchideæ Ordinariæ* (1989), although the synthrumental procedure used is identical. Reproduced by permission of Clarence Barlow.

Poller identifies that, within a broader composition, synthrumentation musically functions as "one component among others," *i.e.* with various degrees of independence from other parameters such as melody, rhythm and form.[11] Accordingly, in *Im Januar am Nil*, using self-generated speech audio as a target, Barlow applies his synthrumental realisations to an arbitrary linear melodic invention (initially in the bass clarinet), with the intent of making it "sound like speech."[12] Regarding this, Barlow writes:

[11] Poller, "Clarence Barlow's Technique of Synthrumentation," 9;
Ibid., 22.
[12] Clarence Barlow quoted in Stephan Kaske, "A Conversation with Clarence Barlow," *Computer Music Journal Vol. 9, No. 1* (1985): 27;

> "It is the string section which is synthrumentally treated with an underlying bass clarinet explicitly but softly playing the melody. The analysed sound material is a set of sentences in the German language excluding all phonemes containing noise spectra such as plosives and fricatives. In all a total of two hundred words were found based on the remaining - lateral, nasal and vowel - phonemes, out of which a number of ,,meaningful" sentences were formed, e.g. *An Müllmänner in Armenien nun ein Jahr lang erinnern* ("Now commemorate garbage collectors in Armenia for one year"[…])."[13] (see Figure 1.2.)

And, separately, in the liner note to the Ensemble Köln recording of *Im Januar* from 2018:

i.e. it is only the sonic or phonemic character of speech that is being attempted to be captured, rather than any aspect of prosody, intonation, cadence, *etc.*

[13] Barlow, "On the Spectral Analysis of Speech," 184.

> "The timbre derives from the synthrumentation of nine concocted German sentences (e.g. one which contains the title itself: *Im Januar am Nil, Mumien* anmalen = "In January at the Nile, painting mummies"). All text syllables are spectrally harmonic, comprised of vowel, approximant, liquid and nasal phonemes. Ideally the bowed "words" should be comprehensible, but an ensemble of seven string instruments can only be approximative."[14]

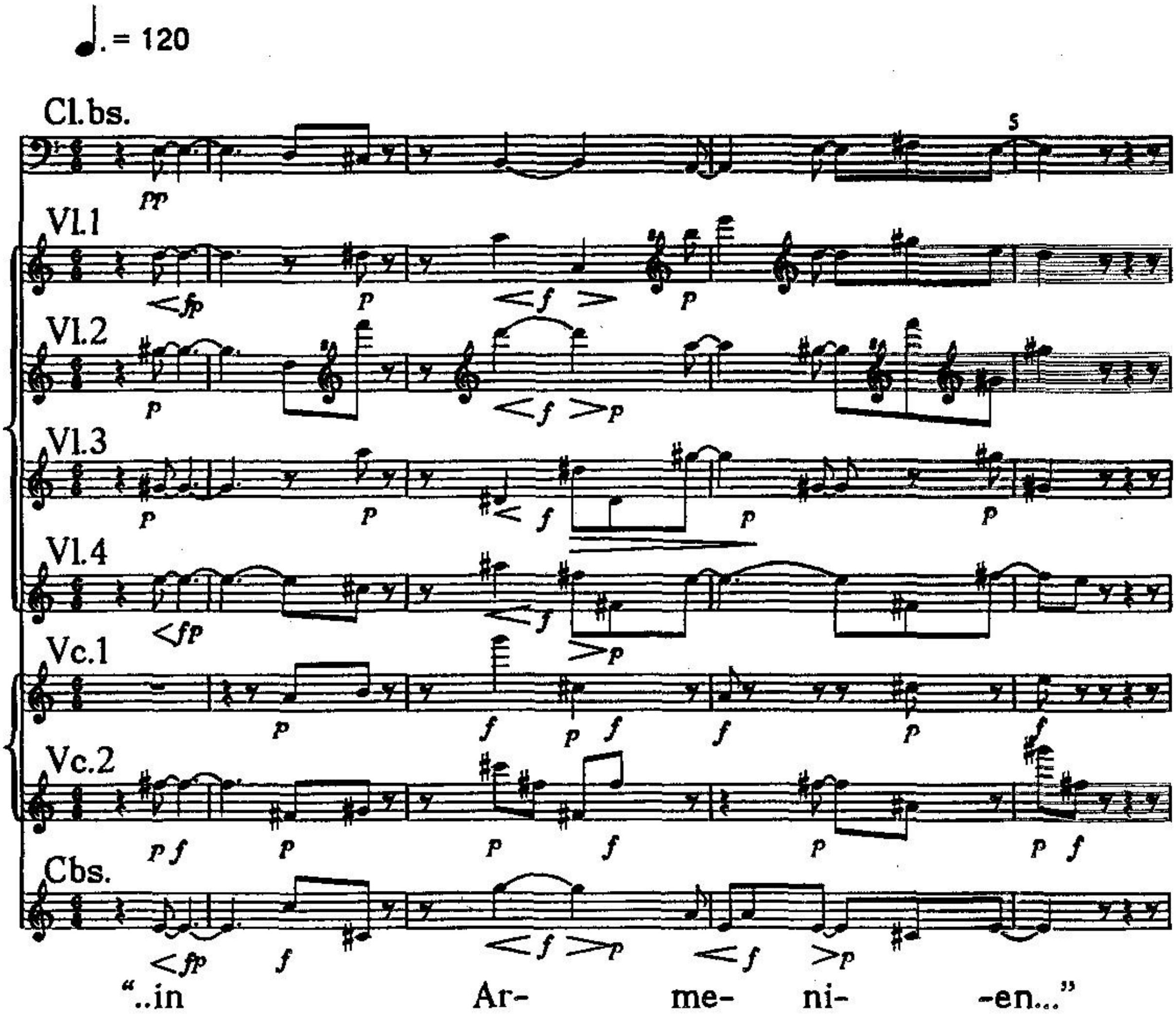



Figure 1.2. Excerpt from *Im Januar am Nil*, as reproduced in Barlow's *On Music Derived from Language* (2009). Reproduced by permission of Clarence Barlow.

These descriptions are illustrative, and their specificity already suggests some aspects of additive instrumental resynthesis that might remain unexplored or overlooked. However, Barlow's primary documentation on the synthrumental technique tends not to elaborate much

[14] Clarence Barlow in liner note to *Musica Algorithmica*, Ensemble Köln, Robert H.P. Platz, Iceland Symphony Orchestra, Hermann Bäumer, Südwestfunkorchester Baden-Baden, Ingo Metzmacher, Ensemble Modelo62, and Ezequiel Menalled (Maria de Alvear World Edition 0034, 2018) 2 CDs, 5-6.

further beyond these base technical aspects – especially regarding the rationale behind certain aspects of procedure, which although implied is not made explicitly clear. As such, prior to any speculations on new methodological start-points, I will first situate *Im Januar* and synthrumentation in the context of other musical works who share similarities of compositional focus or material technique.

1.2. 'Pre-technological' compositional use of externalised vocality

As is the nature of most all musical (let alone artistic) concepts, synthrumentation shares some commonalities with its direct or indirect predecessors in other musical works. A most direct link can be made in the first instance between synthrumentation and other extant orchestrations that intend to identifiably simulate or synthesise discrete-but-corporeally-absent sounds. The simple imitation of external sounds in music is not particularly novel in and of itself, and has been common across human cultures since ancient times – although then primarily in functional contexts, as part of medicinal or spiritual practices, or sympathetic magic.[15] The later emergence of more self-contained, 'artistic' approaches to musical simulation of the external can be identified alongside the development of Western musical practices;[16] and, regarding disembodied human vocality specifically, numerous examples can be found in the late 19th and early 20th centuries: Verdi's *Rigoletto*, Wagner's *Siegfried*, Debussy's *Pelléas et Mélisande*, and Chausson's *Le Roi Arthus*, Holst's *Planets,*

[15] Curt Sachs, The Rise of Music in the Ancient World, East and West (New York: Dover, 2008), 19-23, 47;
D. P. Walker, *Spiritual and Demonic Magic: From Ficino to Campanella* (Pennsylvania: PSU Press, 2003), 25-26;
Michael Thaut, "Music As Therapy in Early History," In *Music, Neurology and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies*, edited by Eckart Altenmüller, Stanley Finger and François Boller (Amsterdam: Elsevier, 2015), 146-149.

[16] Such examples are numerous, and any substantial discussion on this front lies beyond the scope of this paper – as a brief account, works of this description include: in the Classical era, Haydn's *Toy Symphony*; the 19th century, anvils in Verdi's *Il Trovatore* and Berlioz's *March to the Scaffold*; in the 20th, nightingale calls in Resphighi's *I Pini del Gianicolo*, Stravinsky's *Jeu du Rossignol Mécanique*, Mosolov's *Zavod,* and many more; arguably far earlier too in the baroque era, as seen in Rebel's *Le Cahos*. Note too the relatively vast number of years spanned by this practice, particularly when compared to 'vocality-only' attempts as discussed later in this Chapter.

and so on; each of these feature unseen or obscured singers with the explicit intent of evoking "something more than the speech of a human being"[17] – that is to say, such 'externalised' approaches transcend literal depiction, as a matter of artistic intent.

However, more in keeping with *Im Januar*, likely the first significant work to simulate human vocality without the presence of the voice itself is Ravel's *Daphnis et Chloe*, specifically during the *Danse grotesque du Dorcon* of Part One.[18] At rehearsal mark 41, a crowd's laughter is portrayed sequentially across two strata in the orchestra's upper tessitural range, with dissonant, grace-note-laden punctuations in the winds overlaid by fractured tremoli in the upper ranges of the strings. Ravel does this purely by means of acoustic instruments – *even with* a full massed chorus at his disposal for the potential purpose of vocal effects (see Figure 1.3).

[17] Giuseppe Verdi, *Rigoletto* (Milan: Ricordi, 1914), 357-358;
Carolyn Abbate, "Debussy's Phantom Sounds," *Cambridge Opera Journal Vol. 10 No. 1* (1998): 70-73.

[18] I acknowledge that, in a broad sense, orchestrational practices from the Renaissance onwards involving the trombone and vocal music (famously paired for their timbral and intonational similarities) could be argued to constitute 'simulation' to some extent – however I would assert that the trombone in these cases primarily serves to potentiate an existing literal vocal affect, rather than being a self-contained and solely-intentioned mimetic device unto itself; these practices are therefore omitted here.

[IMAGE REMOVED DUE TO COPYRIGHT]

Figure 1.3. Excerpt from *Daphnis et Chloé: Danse grotesque de Dorcon*, rehearsal mark 41-42, from Maurice Ravel, *Daphnis et Chloé* (Paris: Durand, 1913), 52.

Ravel lacked meaningful analytic technological aids at the time he wrote *Daphnis* (let alone spectrograms as in Barlow's case), so his simulation of laughter is necessarily reductive – it is orchestrated only in reference to how he remembered or mentally conceived of his target source, and does not reconstruct human voices with any technical accuracy.[19] Equally, however, this is also a deliberate choice: Ravel's reductive translation of human vocality is not simply just an incomplete act of mimicry or crude synthesis – rather, it is its *reductivity itself* that, in turn, through the various ambiguities and metatextual allusions it conveys, allows for the greater evocation of meaning and affect (or, at least, is intended to so do).[20] Much like prior more-literal invocations of externalised vocality, the musical intent here is a transcendent one, enabled by limitations of process. Taken together, all these examples suggest certain functional implications for the use of abstracted or simplified procedure in mimetic musical contexts, that may be visible also in more orthodox synthrumental works.

1.3. The emergence of the *spectrogram*; Toshiro Mayuzumi and *Campanology I* (1957)

Synthrumentation is defined to some extent by analytical procedures, and so, quite naturally, its historical formalisation in musical compositions itself coincides with the development of new analytical technologies alongside. Particularly notable is the emergence

---

[19] With the advent of recording technologies in the late 19th century it is possible that Ravel may have been able to refer to and replay audio snippets for this purpose, but as per Smirnov and Pchelkina this is quite unlikely; usage of recorded material as a referential basis for the compositional process was almost non-existent practice until the end of the 1920s, and even then remained uncommon until the later maturation of *musique concrete*;

See Andrey Smirnov and Liubo Pchelkin. *Russian Pioneers of Sound Art in the 1920s.* Catalogue of the exhibition "Red Cavalry: Creation and Power in Soviet Russia between 1917 and 1945." Madrid: La Casa Encendida, 2011;

Extant works from *Daphnis* until *Im Januar* that make use of similar (non-technological) vocal simulations likely exist, but they necessarily face the same procedural and technical limitations and therefore need not be discussed here.

[20] Jessie Fillerup, "Purloined Poetics: The Grotesque in the Music of Maurice Ravel," PhD diss., (University of Kansas, 2009), 263-264;

Michael J. Puri, *Ravel the Decadent: Memory, Sublimation, and Desire* (Oxford: Oxford University Press, 2011), 92-93;

Fillerup and Puri's intersecting discussions demonstrate that, whilst there is still much debate amongst scholars on the discrete meanings that Ravel wanted to communicate in his work, there exists a notable, fundamental consensus that – particularly in the case of Ravel's 'laughter' – the invocation of external sounds carries a transcendent intent beyond just dramatic potentiation.

of *spectrograms* – automatically produced graphical depictions of sound – which, although theorised since the 19th century, only fully came to fruition with the technological boom of WWII.[21] Early devices invented for this purpose were all mechanical; it was not until the 1970s that the generative processes for spectrograms became largely digitally-based, and processes such as Fast Fourier Transform (FFT) analysis were in turn more readily available for integration into composers' practice[22] – this is true of the *spectralists*,[23] as well as Barlow himself.

That is not to say, however, that the use of FFT analysis in the composition of music was feature exclusive to works after 1970: notably, Mayuzumi's *Campanology* (1957) for orchestra – putting aside any broader anticipation of *spectralism* in general – is likely the first, fully-fledged example of a synthrumental work.[24] It is the first in a series of pieces by Mayuzumi (not necessarily all involving synthrumentation) invoking spectral methods, and was later repurposed (under the title *Campanology I*) as the first movement of his *Nirvana Symphony* (1958).[25]

The core material of the original *Campanology* is derived from a number of recordings of *bonshō* (a traditionally important type of Japanese temple bell) – some self-generated and some derived from radio broadcasts; Mayuzumi later had their frequency contents analysed, and, invoking the spectrographic methods of an earlier 1948 experimental

---

[21] W. Koenig. et al., "The Sound Spectrograph," *The Journal of the Acoustical Society of America, Vol. 18 No. 1*. (1946): 24.
[22] Himanshu Chaurasiya, "Time-Frequency Representations: Spectrogram, Cochleogram and Correlogram", *Procedia Computer Science Vol. 167* (2020): 1904;
For further discussion on the technical aspects of FFT analysis, see Ch. 1.4. and Ch. 3.1.
[23] See Ch. 1.4.
[24] See definition in Introduction. There is also a reasonable argument to be made that Mayuzumi anticipated *spectralism* by over a decade and renders any arguments over its variously Romanian or French origins moot.
[25] Shimizu, Yoshihiko. "The Creative Quest into Temple Bell Sonorities: Works of *Musique Concrète* by Toshiro Mayuzumi." *Contemporary Music Review Vol. 37, No. 1*, https://doi.org/10.1080/07494467.2018.1453335 (2018): 3.

acoustics paper by Keiji Yamashita, directly extrapolated three primary chords which would form the basis of the work (see Figure 1.4). [26]

[IMAGE REMOVED DUE TO COPYRIGHT]

Figure 1.4. Three *bonshō*-transcription-derived primary chords from Mayuzumi's *Nirvana Symphony*, *i.e.* also *Campanology (1957)*, as replicated from primary source documentation by Takakura (2017).

In spite of the apparent rigour of Mayuzumi's inaugural synthrumental method, there are notable idiosyncrasies present within; chief of these is a wilful reductivity of process in his initial transcriptions of target material. Resonant metal objects, including but not limited to bells, consist of inharmonic overtone spectra, often entirely unrelated to the harmonic series (and by extension any conventional Western tunings, not least 12EDO).[27] At first it appears that Mayuzumi readily recognised this; he was in fact was attracted to bonshō for their "extremely complex overtone structure, in which […] overtone vibrations [were] not an integral multiple of the number of fundamental frequencies," *i.e.* for their inharmonicity;[28] in

[26] Yuriko Takakura, "A Comparison of the Compositional Process between the *Nirvana Symphony* and the *Mandala Symphony*: An Analysis of the "Campanology Documents," *Ongakugaku, Vol. 63, No. 2* (2017): 64-65;

Takakura, "A Visual Analysis Methodology for Music Compositional Process with Sound Resynthesis," PhD diss. (Keio University, 2019), 29.

[27] The development of both historic Western temperaments as well as the more modern 12-tone equal temperament was heavily influenced by intervals (and perceived importance thereof) present in the harmonic series; this is particularly well exemplified in Pythagorean tunings (based on the frequency ratio 3:2) and Meantone tunings (based on 5:4);

12EDO, that is, '12 equal divisions of the octave' (*a.k.a.* 12-equal temperament, equal temperament, etc.).

[28] Mayuzumi quoted in Judith Ann Herd, "The Neonationalist Movement: Origins of Japanese Contemporary Music," *Perspectives of New Music Vol. 27, No. 2* (1989): 137.

initial sketches he transcribed these inharmonic overtones microtonally (albeit imprecisely).[29] Nonetheless, however, in the writing of *Campanology* itself, Mayuzumi ultimately opted to realise the synthrumented *bonshō* by approximate means (in 12EDO), rendering much of their unique sonic character moot. This reduction is also observable in his later synthrumental work, *Mandala Symphony* (1960), which was based on identical spectral material.[30]

Likewise, it is noteworthy that Mayuzumi's employment of *bonshō* spectra within his synthrumental works lacks the integrative nuance later associated with the technique. In *Campanology* (1957), bell resonances are employed more as pure, static, repeated objects, rather than base materials to themselves be manipulated.[31] It has been argued that the literalism of Mayuzumi's orchestrations does not in itself translate to a literalism of intent, however; Shimizu remarks that Mayuzumi's particular treatment of spectral sources in *Nirvana Symphony* (and *Campanology* by extension) "can be interpreted as [an] attempt to solidly engrave characteristics of 'Japan-ness'"[sic][32] into his work – a remark that implies an apparent unique communicative potential to the limitations present within Mayuzumi's method.

Mayuzumi ended up producing a number of works under the 'Campanology' banner, for a variety of forces, from 1957 to 1967;[33] not all of these involved synthrumental practices, however, and, ultimately, his use of the technique largely subsided after *Mandala Symphony*.[34] Particularly in light of the composer's budding nationalistic political leanings (which were later a significant component of his mature output), Shimizu suggests that these purely acoustic works – and perhaps the particulars of procedure associated with them – were

---

29 Takakura, "A Visual Analysis Methodology for Music Compositional Process with Sound Resynthesis," 34.
30 *Ibid.,* 51.
31 *Ibid.,* 29-30.
32 Shimizu, "Creative Quest," 11.
33 Shimizu, "Creative Quest," 3.
34 Yoshihiko Shimizu, "Serial Technique in Toshiro Mayuzumi's 'Campanology Effect'," *Journal of the Musicological Society of Japan, Vol. 56, No. 1* (2010): 26.

procedurally incongruous with Mayuzumi's communicative goals, and he found greater possibility in the purely acousmatic *musique concrète* in this regard.[35] Given the early death of Japanese synthrumental practice, no major movement was able to follow in *Campanology*'s wake – much unlike developments that were soon to take place in Europe, a decade later, namely in the form of *spectralism*.

1.4. Synthrumentation and French spectralists Gérard Grisey and Tristan Murail

The advent of spectralism in the 1970s marked a significant turn in Western music: not only towards a new focus on "overall control of the musical spectrum,"[36] a unique, 'matter-first' perspective on form,[37] and ignition of interest in "timbre as a fundamental factor in composition,"[38] but also for the popularisation of employing sonic analysis as an aid for the creation of acoustic music. Unsurprisingly, the practice of synthrumentation and spectralist aesthetics overlap heavily; synthrumentation itself is identifiable in numerous pieces written prior to *Im Januar,* which in all likelihood themselves served to some degree as musical influences (direct or indirect) in Barlow's work.[39]

Unfortunately, it is not feasible here to exhaustively chart every single appearance of synthrumentation throughout spectralism as a whole; instead I will opt focus on a number of specific works that I feel best exemplify the coincidence of these practices, and are most salient to this paper's aims. From my research, synthrumentation's earliest appearance in spectralist composition was in in Gérard Grisey's *Partiels* (1976); synthrumentations appear to be limited to the compositional output of French spectralists, at least in this early period;

---

[35] Shimizu, "Creative Quest," 11.
[36] Hugues Dufourt*, "*Musique Spectrale," in *Musique, pouvoir, e´criture* (Paris: Christian Bourgois, 1991), 291. Quote trans. Féron in "The Emergence of Spectra."
[37] *Ibid.*
[38] Viviana Moscovich, "French Spectral Music: An Introduction," *Tempo, New Series, No. 200* (1997): 21.
[39] Although scholars have forwarded slightly differing understandings (and chronologies) of what *spectralism* entails, the association of some specific composers with the movement is a matter of strong consensus; see Bibliography.

finally, there exists a common scholarly understanding that, of the French spectralists, "Gérard Grisey and Tristan Murail are the two best-known" and arguably most foundational and influential upon both their contemporaries and later adherents;[40] hence, discussions herein will be limited to exemplary works of Grisey and Murail, accordingly.[41]

1.5. Synthrumentation in Grisey's *Partiels* (1976)

Grisey's *Partiels* (1976) from the cycle *Les Espaces Acoustiques* (1974 - 1985) is perhaps the most iconic and well-documented French spectralist work,[42] and one so influential and demonstrative that it has been described as a "sort of manifesto of 'spectral' aesthetics" in and of itself.[43] Grisey himself had conceived of "instrumental synthesis […] modelled on the principles of (electronic) additive synthesis" some years prior as documented in an article from 1973,[44] and *Partiels* – as a work derived at the outset from the spectrogram of a trombone playing the pitch E_2 [45] – was his first fully-fledged realisation of this concept, and is likewise a complete synthrumental work.[46]

---

[40] François Rose, "Introduction to the Pitch Organization of French Spectral Music," *Perspectives of New Music Vol. 34, No. 2* (1996): 6.

[41] N.B. the chosen limitation to Grisey and Murail here is a resultant matter of this paper's specific research aims, rather than an erasure of the contributions of others to the movement; I acknowledge that the debate over spectralism's geographical or national origins remains contentious, particularly regarding competing 'spectralist' ownership claims by French and Romanian composers (particularly Hughes Dufourt and Horațiu Rădulescu, respectively) and their adherents. Putting aside the fact that Mayuzumi's spectrographic work essentially renders this topic a moot point more than a decade in advance, Rădulescu's primary claim centres around his *Credo* of 1969/76, which is explicitly *not* a work that invokes synthrumental techniques, and therefore is not included in discussions.

[42] Liner note to Gérard Grisey, *Les Espaces Acoustiques*, Garth Knox, Asko|Schönberg, WDR Sinfonieorchester Köln, and Stefan Asbury, (KAIROS), 2005 CD 0012422KAI;

There exists conflicting documentation regarding the exact year in which *Partiels* was completed.

[43] François-Xavier Féron, "Gérard Grisey: première section de *Partiels* (1975)," *Genesis, Vol. 31* (2010): 77.

[44] *Ibid.*, 79; translation my own.

[45] Joshua Fineberg, "Musical Examples," *Contemporary Music Review Vol. 19, No. 2* (2000): 115-117.

[46] It would be remiss not to note that the ending of the previous work in the *Les Espaces* cycle, *Périodes*, actually makes use of the exact same synthrumental derivation as a kind of coda and bridging material between the two pieces; of course, however, its usage compared to as in *Partiels* is less substantial.

Notable in-depth analyses of *Partiels* have been produced by Féron and Krier;[47] in a simple, macrostructural sense, the work can be described as a series of orchestrated realisations of a trombone spectrum, gradually distorted and irregularised though changes of instrumental technique (and resulting timbres), as well as the introduction of other, 'synthetic' pitch materials.[48]

At the outset, Grisey's application of synthrumental processes is quite simple, and resembles that of Mayuzumi's (in that it literally replicates the transcribed overtone content of a spectrogram); after a stark opening with the trombone playing a forte E_2, and reverberations in the double bass an octave below, the other instruments, each fixed to a pitch within that same note's harmonic series, slowly emerge in a cloud of resonance (see Figure 1.5.). In a notable shift from previously cited works, Grisey notates microtonal pitches into the score, allowing for more faithful replication of existing spectra.

[47] See Féron (2010) and Yves Krier, "Partiels, de Gérard Grisey, Manifestation D'une Nouvelle Esthétique." *Musurgia Vol. 7 No. 3-4* (2000): 145-172;

Féron's analysis notably makes reference to primary source pre-compositional planning materials left by Grisey.

[48] Krier, 156-170.

[IMAGE REMOVED DUE TO COPYRIGHT]

Figure 1.5. Opening page of Grisey's *Partiels* (pub. 1976).

What is particularly of interest here, however, is Grisey's chosen point of divergence from the basic aspects of synthrumental practice; as the work progresses, new pitches not found in the E_2 harmonic spectrum are introduced, through both the reification of (non-harmonic) difference tones,[49] and the subsequent emergence of what Krier describes as 'halos' – essentially, reinforced high partials from said new pitches (see Figs. 1.6. and 1.7.).[50] In a sense, Grisey uses his initial spectrum to derive new pitch materials, which are then treated as new theoretical fundamentals unto themselves.[51]

---

[49] Féron (2010), 86; Krier 162-163;

Difference tones are an acoustic phenomena that result from the coincidence of two pitches sounded at the same time; their sounding frequency is literally the difference between the frequencies of said two pitches. As both Féron and Krier note, difference tones are usually only faintly audible, if at all, save for literally realising them in orchestration as Grisey has chosen to do so here.

[50] Krier, 162-163.

[51] It is worth noting the procedural similarity here with the pitch and scale derivations of the earlier Erv Wilson and Harry Partch, albeit in explicitly non-representational contexts.

[IMAGE REMOVED DUE TO COPYRIGHT]

Figure 1.6. Krier's diagram of 'halation'.

[IMAGE REMOVED DUE TO COPYRIGHT]

Figure 1.7. Corresponding section from *Partiels*, trimmed from pages 18-19.

Grisey describes this emergent process in the piece as a "natural spectrum [drifting] with each repetition towards inharmonicity," and likewise, both Krier and Féron discuss notions of the musical fabric becoming more (and occasionally, less,) 'inharmonic'.[52] In a strict sense this is of course true; Grisey at the outset establishes clear boundaries of what constitutes 'harmonic' pitch material, via his spectral analyses; any pitch content outside these boundaries (both discrete, as well as timbrally resultant from technique changes such as *sul ponticello*) can be viewed as non-harmonic as a result.

An important observation, however: inharmonicity in *Partiels* is an emergent property of a *strictly harmonic concept*, rather than a fundamental point of departure unto itself. Grisey's basic material, and by extension his synthrumental technique, is entirely bound to the harmonic series; the appearance of new pitches is necessarily a result of non-synthrumental processes.

1.6. Inharmonicity and its intrinsic role in sound (an aside)

It would be tempting to conclude, from the reductivities present in Grisey's process above, as well as in Barlow's *Im Januar,* that inharmonicity is a structurally unimportant component of sound, or at least does not warrant particular attention in the additive instrumental resynthesis process. Inharmonicity is in fact, however, a ubiquitous component in the mechanics of sound generation, as well as resultant sound itself, and this has tended to be obfuscated or otherwise overlooked in existing musical scholarship and compositional practice, as above.

[52] Féron (2010), 93;
Krier, 162-170.

A harmonic sound is often understood to be one whose partials are fully in keeping with the ratios of the mathematical harmonic series,[53] and is suggested to be an intrinsic result of *periodic* modes of sound production ("characterised by repetitive structure," as in the vibration of the larynx when speaking, the bowing of a string, or the buzzing of the lips on a trombone);[54] accordingly, even periodic input into a non-harmonic resonator will produce a perfectly harmonic sounding result, in an acoustic phenomenon termed *mode locking*.[55] Conversely, under this definition, inharmonic sounds might be those whose mode of generation are non-periodic, such as the momentary striking of a drum, plucking of a string, or rustling of the wind.

This understanding is not completely reflective of physical reality, however; outside of purely electronic contexts, the appearance of harmonicity and inharmonicity in a given sound source tend to be interlinked. For one, inharmonicity is present in the transients of sounds (in musical terms, a note's articulation), which are themselves a vital component of a given instrument or source's identifiable character;[56] furthermore, it pervades the sustained portion of many (ostensibly 'harmonic') periodic sounds, in the form of a 'residual' or noise component:[57] these residuals are "often important to the integrity of [a musical instrument's] signal" and its subsequent perception.[58]

---

[53] This is certainly true of Grisey in *Partiels,* Barlow in *Im Januar,* and later examples to be discussed.

[54] Peter M. C. Harrison and Marcus T. Pearce, "Simultaneous Consonance in Music Perception and Composition," *Psychological Review Vol. 127, No. 2* (2020): 217.

[55] Wolfe, Joe. "How Harmonic are Harmonics?" *University of New South Wales School of Physics*. Accessed May 21, 2021. https://newt.phys.unsw.edu.au/jw/harmonics.html;

N. H. Fletcher, "Mode Locking in Nonlinearly Excited Inharmonic Musical Oscillators," *The Journal of the Acoustical Society of America Vol. 64, No. 6* (1978): 1566.

[56] Mark Every, "Separating Harmonic and Inharmonic Note Content From Real Mono Recordings," *University of York, Department of Electronics* (2005): 1-5. https://www-users.york.ac.uk/~jes1/EveryDMRN05.pdf;

Stephen McAdams and Bruno L. Giordano, *The Perception of Musical Timbre,* edited by Susan Hallam, Ian Cross, and Michael Thaut (Oxford: Oxford University Press, 2012), 4-5.

[57] Every, 1;

Tony S. Verma and Teresa H. Y. Meng, "Extending Spectral Modeling Synthesis with Transient Modeling Synthesis," *Computer Music Journal Vol. 24, No. 2* (2000): 54-58.

[58] *Ibid.,* 47.

Residuals culminate from many acoustic factors; notably, for musical instruments, they are often linked to the "always non-linear," non-periodic character of human input (bowing, buzzing, *etc.*) prior to the occurrence of mode locking:[59] for example, the interaction between bow and string on a violin and resulting wispy noise content separate from musical pitch. The inherent inharmonicities of most physical resonators can also occasionally interfere with the function of mode locking at extremes of amplitude or frequency, resulting in harmonically destabilised sonic output; as an aside, mode locking, often framed as an enabler of instrumental harmonicity, ironically is itself the cause of the phenomenon of (inharmonic) multiphonics.[60]

Importantly: all of the above is true of the audio targets resynthesised in the works discussed: *e.g.* the trombone,[61] which in Grisey's working practice has been simplified in terms of perfect, idealised harmonic partials; or, the human voice – reduced to harmonic partials by Barlow, and whose inharmonic character will be discussed further in Section 2.

1.7. Synthrumentation and treatments of inharmonicity in Grisey's *Modulations* (1976-77), and Murail's *Désintégrations* (1982), *L'Esprit des dunes* (1994), and *Gondwana* (1980)

This reductive harmonicity is present in Grisey's later *Modulations* (1976-77), also from the same cycle. The initial material of the piece – another example of synthrumentation

---

[59] N. H. Fletcher, "The Nonlinear Physics of Musical Instruments," *Reports on Progress in Physics Vol. 62, No. 5* (1999): 727-731.

[60] Fletcher, "Mode Locking", 1568;

Alistair C. P. Braden, Michael J. Newton, and D. Murray Campbell, "Trombone Bore Optimization Based on Input Impedance Targets," *The Journal of the Acoustical Society of America Vol. 125, No. 4* (2009): 2406.

[61] Braden et al., 2404-2412.

– is directly derived from the spectra of variously muted trombones; yet, in spite of changes in the composer's demonstrably rigorous analytical outlook and his written acknowledgement of the inharmonic components and potentials inherent to these sounds (in his words, "formants, passing notes, differential tones, white noise, filtering, *etc.*"),[62] the work's synthrumental derivations are wholly harmonic, *i.e.* the piece still ultimately sees inharmonicity as an emergent, non-generating property.[63]

This is similarly true of much of Murail's apparently synthrumental works: the orchestrations in *Désintégrations* (1982) make direct use of the harmonic content of various spectra (that of "low piano notes, brass instruments, and the cello"), with inharmonicity solely resulting from later interventions;[64] the more 'purely instrumental' sections of *L'Esprit des dunes* (1994) involve the spectra of both human voices and Tibetan trumpets, but, in any case, the blending of the harmonic with the inharmonic (*e.g.* the "sounds of tearing paper with instrumental timbres") is left to electronic realisations.[65] Inharmonicity is still largely absent from any purely acoustic orchestration in these works.

As a notable exception, Murail takes a mixed approach in *Gondwana* (1980) for orchestra, treating inharmonicity both as a starting point and emergent property. The work explicitly seeks to emulate inharmonic, "bell-like" spectra, which are derived in the first instance from the frequency modulation of two arbitrary pitches (G_4 and $G\#_3$).[66] So too, unlike in Grisey's *Partiels*, Murail makes use of multiple generative pitch sets, and

---

[62] Gérard Grisey, liner note to *Les Espaces Acoustiques*. Garth Knox, Asko|Schönberg, WDR Sinfonieorchester Köln, and Stefan Asbury (KAIROS 0012422KAI, 2005), 2 CDs.
[63] Rose, 35-36.
[64] Tristan Murail, "Spectra and Pixies," *Contemporary Music Review Vol 1, No. 1* (1984): 161-162;
Not to mention the piece involves fixed electronics, too, which somewhat throws its categorisation as 'synthrumental' into doubt.
[65] Anderson in liner note to *Tristan Murail: Serendib / L'esprit des dunes / Désintégrations*, Ensemble Intercontemporain, David Robertson (Accord AC4653052, 1996), 1 CD.
[66] Rose, 30-31;
Tristan Murail, "Gondwana," *Tristan Murail*, accessed May 16, 2021, https://www.tristanmurail.com/en/oeuvre-fiche.php?cotage=TR1572.

throughout the work interpolates the above with separate, harmonic, overtone-derived chords.[67]

However, although the techniques used here are similar to synthrumental works discussed previously, it is difficult to definitively accept *Gondwana* into the category. Murail does, granted, make systematic use of certain 'real-world'-derived sonic phenomena, such as the non-uniform decay of bell sounds;[68] in spite of this, though, all chordal materials in *Gondwana* are ultimately self-invented – they are not sourced directly from any spectral models, and are instead entirely synthetic.

It is apparent that exclusive adherence to harmonicity seems to pervade common synthrumental practice; Rose rightly observes that, overall, Grisey and Murail, having spearheaded synthrumentation within the spectralist movement proper, "have established the overtone series as their point of reference;"[69] the inharmonic components of spectral sources is largely discarded in the first instance.

Of course, these omissions by no means have any bearing on the works' musical effectiveness – they are, simply put, creative choices. Furthermore, as Anderson notes, "the use of spectra […] is only the most superficial feature of the music of these composers;"[70] much in keeping with other synthrumental and adjacent works discussed, Grisey and Murail's conceptual intentions were *not* to hyper-realistically emulate recorded sounds; they instead wanted to conjure a transcendent experience, and sonic expression of their respective creative philosophies. Still, however, from a processional point of view, and in the context of this

---

[67] Rose, 34.
[68] *Ibid.*, 32.
[69] *Ibid.*, 6;

Again, *synthrumentation* is used here as a general category.
[70] Anderson, "A Provisional History of Spectral Music," 7.

paper, the absence of inharmonic sources in most all synthrumental writing is conspicuous; it necessarily invites further attention.

1.8. Mixed acoustic-electronic perspectives on speech synthesis: Jonathan Harvey's *Speakings* (2008) and Peter Ablinger's *phonorealist* works

Before reviewing all prior discussions, it is worth at this point acknowledging two other particular musical examples – both electro-acoustic, rather than purely acoustic – that necessarily do not employ synthrumentation, yet maintain strong procedural and aesthetic commonalities with it. Namely, these are Jonathan Harvey's *Speakings* (2008) for orchestra and live electronics, and Peter Ablinger's ongoing series of 'phonorealist' speech-synthesis works for player piano (amongst other forces); both may provide unique analytical perspectives with relation to speech synthesis, and potential comparisons that could be made to Barlow and others' working methods as a result.

Speakings was possibly the most ambitious and innovative of Jonathan Harvey's late large-scale works: it served as a singular intersection between his own mature acoustic practice, electronic practice, and Buddhist philosophical outlook; it was also at the same time a highly collaborative and logistically complex work, facilitated in significant part by Gilbert Nouno and IRCAM. As implied by its title, the piece explores the possibility of 'making an orchestra speak,' letting "vowel and consonant spectra-shapes flicker […] across orchestral textures;" [71] in Harvey's words, "[this] process of 'shape vocoding' […] is the main idea of this work." [72] Nouno later published a technical paper detailing some of the processes involved in *Speakings*' treatment of speech.[73]

---

[71] Jonathan Harvey, *Speakings* (London: Faber, 2008), iv.
[72] *Ibid.*
[73] Gilbert Nouno, Arshia Cont, Grégoire Carpentier, and Jonathan Harvey, "Making an Orchestra Speak," *Sound and Music Computing* (2009): 1-6. https://hal.inria fr/hal-00839067/document.

Broadly, there are two contrasting modes by which the orchestra becomes speech-inflected: by acoustic (or 'orchestrational') means, and, by live-electronic means. In the case of the former category, Harvey makes use of IRCAM's proprietary *Orchidée* computer-assisted orchestration program to, "given an input target sound, [output] a musical score for imitating [that] sound;"[74] this musical score (or orchestration) is then performed by live players without any electronic intervention.

The software seems on its surface to be an ultimate, automated method for synthrumentation; however, it is still limited by procedural compromise: as explained in *Orchidée*'s own technical paper, the assisted orchestration process consists of a number of intermediary steps before a score is actually outputted. Particularly relevant is the automated generation of an internal 'SoundTarget' object from an audio input, which, amongst other things, reduces the spectral content of said input to a set of discrete pitches (by way of peak analysis) approximated in 24EDO pitch-space (*i.e.* quartertones); the user is then provided a visual representation of the SoundTarget, and given the option to alter it by way of harmonic overtone reinforcement or application of other filters, before moving to later steps;[75] finally, the user is suggested a number of orchestration solutions generated from the reduced SoundTarget, and can then refine their choices with specific reference to resultant timbre (from a finite set of possible orchestral sounds and techniques).[76]

Orchidée is highly sophisticated; its applications to speech are broader in scope than Barlow's original technique, and it does not innately limit target material choice in the first instance (unlike, say, Barlow's preference for vowels over consonants). However, much of its

[74] *Orchidée* has since largely been succeeded by IRCAM's very recent, also-proprietary *Orchidea*, a similar software solution for assisted orchestration; although significant differences exist between them, direct comparisons are not well documented, so inevitably some discussion here may be out of date at time of publication.
[75] Grégoire Carpentier and Jean Bresson, "Interacting with Symbol, Sound, and Feature Spaces in Orchidée, a Computer-Aided Orchestration Environment," *Computer Music Journal Vol. 34, No. 1* (2010): 17-22.
[76] *Ibid.*, 21-22.

strengths lie in the artistic flexibility it offers users *after the fact* of its initial, rudimentary synthrumental suggestions, which are themselves derived in a procedure highly similar to composers' practices previously discussed; they generally aim to describe a target sound in terms of discrete pitch content. There also exists an apparent emphasis on harmonicity within the manually-controlled aspects of the software, suggesting somewhat a continuation of previous analytical blindspots.

Quite illustratively, in *Speakings*, it is the live-electronic (*i.e.* non-*Orchidée*-generated) component that explicitly attempts to compensate for this, and in so doing pay greater attention to the inharmonicity of speech. Nouno notes that "there seem to be a lot of interesting structural information[sic] that are seemingly lost in [the assisted orchestration process],"[77] and, to compensate, the work calls for a small ensemble group to be separated from the main orchestra and affixed by specific microphone placement. This raw audio is then live-processed such that "the inner-rhythmical [and dynamic formant] structures of speech are stamped into the [sound];"[78] the result is further diffused in a specific surround-sound speaker configuration around the audience.[79] Ultimately, electronic intervention is necessary here to fully realise a balanced representation of speech; for all its scope and technical achievement as an overall work, *Speakings*' purely acoustic models, when viewed in isolation, are constrained in much the same way as previous synthrumental pieces.

Even further afield, Ablinger's 'phonorealist' pieces, variously written since 1996[80] – which, like synthrumental works, are situated upon acoustic resynthesis and carry a broader intent to reify the "border area between abstract musical structure and [linguistic

[77] Nouno et al., "Making an Orchestra Speak," 4.
[78] *Ibid.*
[79] Jérémie Henrot, "Speakings – Jonathan Harvey: Technical Rider for Performances and Rehearsals," accessed May 23, 2021, https://brahms.ircam.fr/media/uploads/Tech_Rider_-_Generic_for_Speakings.pdf.
[80] Winfried Ritsch, Peter Ablinger, and Thomas Musil, "Robotic Piano Player Making Pianos Talk," *Institute for Electronic Music and Acoustics Graz* (2011), 1.

cognition]"[81] – tend to completely exclude the presence of live players, even if the works are themselves for acoustic instruments: for example, *Deus Cantando* (2009) for player piano, which resynthesises a young German speaker's recitation of a declaration from the International Environmental Criminal Court.[82] In this work, the compositional process similarly starts with digital audio input and subsequent spectral analysis,[83] but procedurally diverges rather significantly from typical synthrumental practice. When working with a given target, Ablinger, rather than prioritising dominant pitches or harmonic peaks, instead accounts for as much of the spectrum of an audio target as possible in his analysis – by a measured combination of FFT, constant-Q transform (CQT), and wavelet analysis (all dissimilar techniques).[84] The subsequent resynthesis process is then a fairly simple translation of precisely derived velocities, durations and pitches (at a given resolution) to MIDI information, approximated in the standard 12EDO of a piano.[85]

The sonic results of Ablinger's technique are immediate and incredibly striking, and themselves have been able to make a somewhat significant cultural impact in recent years; subsequent, derivative works have seen widespread distribution across popular internet platforms, and occasional intersections with meme culture,[86] implying a unique effectiveness of the technique as a communicative artistic tool. Unfortunately, however, the methods of the 'speaking piano' generally aren't translatable to 'normal' synthrumental contexts: because of Ablinger's reliance on strict parametric precision to produce his desired effect, faithful replication by a live, human ensemble would be nigh impossible, much like the predicament

[81] Peter Ablinger, "Phonorealism." Accessed June 1, 2021. https://ablinger.mur.at/phonorealism.html.
[82] G. Douglas Barrett, *Window Piece: Seeing and Hearing the Music of Peter Ablinger* (self-published), 5.
[83] *Ibid.*
[84] Ritsch, "Robotic Piano Player," 2;
See Pulkki and Karjalainen, *Communication Acoustics*, 53, for a brief dissection of what FFT, CQT and wavelet analysis involve.
[85] Ritsch, "Robotic Piano Player," 2-3.
[86] See "Auditory Illusions: Hearing Lyrics Where There Are None," https://www.youtube.com/watch?v=ZY6h3pKqYI0; and "Shrek but the ENTIRE MOVIE is converted to MIDI," https://www.youtube.com/watch?v=wcehaxidJZk.

faced by Barlow when streamlining his own original process.[87] However, *Deus Cantando* and its ilk still suggest a way forward for new approaches to synthrumentation, particularly in terms of analytical scope, and inclusion of wider spectral bandwidths than just those of harmonic peaks.

1.9. *Im Januar am Nil* in context

Given all prior context – particularly discussions over intentional reductivity of process, and inherent limitations of other composers' methods – I want to now draw attention to the following aspects of Barlow's explanations of synthrumentation in *Im Januar* (see long-form quotes in 1.1):

1. His intentional omission of 'noise spectra' from audio sources;

2. The procedure's realisation of (continuous) spectrographic frequency contents in terms of discrete, finite pitch classes (*i.e.* the pitches of 12EDO as shown in the MIDI transcription of Fig. 1.1., and further corroborated by Poller);[88]

3. His acknowledgment of the 'approximative' nature of the additive synthesis' outcome.

With regards to 1. and 2. – much like his spectral and proto-spectral counterparts before him, Barlow's treatment of his source material, *i.e.* the sounds of human speech – is procedurally reductive. In preparing materials for *Im Januar*, he intentionally selected only words containing vowels and nasal phonemes (so as to reduce the initial amount of 'noise' or

[87] I have not mentioned this, for the sake of brevity, but it is worth mentioning that Barlow's prior 'Spectastics' (spectral stochastics) works from the late 1980s onwards do in fact bear some technical resemblance to Ablinger's 'phonorealist' pieces, and do in fact involve live performers; however, these seem to have been realised at lower resolutions than *Deus Cantando* and similar (presumably due to technological limitations), and it is difficult to make any meaningful assessment of Spectastics here in context.
[88] Poller, "Clarence Barlow's Technique of 'Synthrumentation,'" 13.

residual inharmonic content to contend with); Barlow further opted to transcribe (and subsequently generate orchestrations from) only the harmonic content of the resultant spectrograms, discarding any other present inharmonic content in the process.[89]

Barlow further chooses to be approximative in his realisation of pitch: both initially, in employing 12EDO for computer-aided transcriptions; and later, when realising (inherently 'microtonal') higher partials through precise-but-inaccurate microtonal scordatura for individual string parts in offsets of 10, 30, or 40 cents.[90] Ultimately, too, any such use of microtonality as represented within score and parts is still realised wholly within the standard 12EDO notational framework, in a similar fashion to Mayuzumi.[91]

With regards to 3. – the composer's documentation highlights his own conscious acknowledgment of the strengths and limitations of the synthrumental method in fulfilling its intent of making speech recognisable within the musical fabric.[92] However, Poller asserts, in retrospect, that the limitations of Barlow's method only serve to heighten musical and integrative potential[93] – much like in all other composers' work examples discussed.

As such, the reductive or 'approximative' elements of Barlow's methods as identified in points 1. and 2. must be viewed as well-rationalised choices in both a practical and artistic sense: for example, with respect to practicality, the omission of non-nasal consonants in initial material – which, on one hand, readily enables the transcription of harmonic overtones (which to Barlow were ostensibly the most crucial aspect of his speech samples), and on the other, also best enables orchestration for a live, human ensemble (as inclusion of other sonic features of speech may have been impractical and of detriment to overall effect).[94] With

[89] Kaske, "A Conversation with Clarence Barlow," 27.
[90] Poller, "Clarence Barlow's Technique of 'Synthrumentation,'" 13-15.
[91] *Ibid.*
[92] Barlow quoted in Kaske, 27.
[93] Poller, "Clarence Barlow's Technique of 'Synthrumentation,'" 22.
[94] Kaske, 27.

reference to artistic potentiation, Poller notes in addition to his assertions above that Barlow's choices allow "[recognisable aspects of speech] to be completely subservient to the musical composition" as a whole,[95] and grant a notable breadth of flexibility and creative freedom in spite of the apparent rigours of the initial synthrumentation process. This same flexibility is observable in the extrapolative methods of Grisey and Murail; conversely, the pitfalls of methodical rigidity (particularly with regards to musical effectiveness) can be seen in Mayuzumi,[96] and one might speculate too that his eventual abandonment of synthrumental practices is linked.

Importantly, Barlow's specific method is not absolutely fixed from work to work within his oeuvre: it is stripped back even further in the later synthrumental *Orchideæ Ordinariæ* (1989), which "[avoids] melismas and [allocates] specific partials to certain instruments."[97] For the composer, continued narrowing of both procedure and initial target material (for resynthesis) serves only to increase the creative possibility and perhaps even effectiveness of his music. It bears repeating the commonality that exists across all composers discussed (harking back even to Ravel or further): willful reductivity of process – or, ambiguity – in the act of synthesising or mimicking existing sonic objects, is intended to allow their music to communicatively *transcend* the base particulars of said sonic object in and of itself.

It follows that, when developing a new synthrumental method in some way divergent from existing practices, as is the intent of this paper, one must strip away at a target material and hone in on particular aspects of it, rather than trying to directly and pedantically reproduce it as a whole. Simultaneously, said method, at least at the stage of application in

---

[95] Poller, "Clarence Barlow's Technique of 'Synthrumentation,'" 9-13.
[96] Shimizu, "Creative Quest," 11.
[97] Barlow, "On the Spectral Analysis of Speech," 186.

compositional writing, should maintain a degree of flexibility for effective integration into musical contexts (as per Poller), with the overall aim of potentiating both a specific ‘synthesised’ affect *and* overall musical substance in equal measure.

It is clear that historical synthrumental practice demonstrates an overarching bias towards harmonicity, particularly in pre-compositional procedure. Harmonic overtones, or other discrete spectral peaks, seem in most cases to be the sole target or entry point of resynthesis, and have been favoured over other available sonic information within a given audio source. Conversely, it stands to reason that new methods of synthrumentation could be revealed via the conscious invocation of alternate analytical perspectives (such as those evident in parts of Harvey and Ablinger’s work), and through the adoption of a renewed focus on the inherent inharmonicities of sound – particularly, in the context of this paper, the inharmonicities of speech.

CHAPTER 2.

SCIENTIFIC PERSPECTIVES ON SPEECH PRODUCTION

A new method of vocal synthrumentation should be developed with respect not only to extant practices, but also to non-musical scholarly perspectives on speech production and perception. To put it more directly: having investigated synthrumentation from the perspective of synthesis, that is, from musical works after the fact, we ought also to investigate it from the perspective of the target material – speech – itself.

Human speech is complex, and, as with all science, our understanding of it is necessarily limited and constantly undergoing revision. So too, within linguistics and related fields, scholarly interest in the phenomenon of speech production has historically been somewhat sidelined in favour of studies on grammar and semantics; much of more recent and more interdisciplinary research contradicts the understandings of early scholarship, and is likewise varied in the new propositions it puts forward in their place.[98] That is not to say, of course, that consensus is non-existent in the study of speech production, but I do want to make clear at this point that the study of speech production is in many ways still an emerging subfield;[99] it is impossible to provide a balanced account of all scholarly perspectives here as a matter of both practicality and scope – this being an ostensibly music-focused paper.

This is where prior discussions on synthrumentation become quite useful: the apparent limitations of extant practices, orchestrational considerations, and other factors endemic to the genre, can all serve as constraints or reference points in leading science-

[98] Melissa A. Redford, *The Handbook of Speech Production* (Chichester, West Sussex: Wiley-Blackwell, 2015), 2-4.
[99] Philip Lieberman and Sheila Blumstein, *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge: Cambridge University Press, 1988), xiv.

focussed discussions herein. This in turn also allows, in part, for the simultaneous exploration of how new synthrumental methodologies might be developed.

2.1. Components of speech

The communicative phenomenon of speech involves the physical coordination – often automatic or subconscious – of a number of differing bodily mechanisms; as Pulkki and Karjalainen put it, "[speech] is so self-evident an ability that we often don't notice how complex and delicate it is until something goes wrong with it."[100] Lieberman and Blumstein, as a matter of functional convenience, broadly list "three physiological components of speech production": the *subglottal* component, consisting of "lungs and associated respiratory musculature" who regulate breath and generate air flow; the *larynx*, which is a "[generative] 'source' of acoustic energy," responsible for *phonation* (or, in more 'musical' terms, vocal pitch production) and finally, the *supralaryngeal vocal tract*, that is, "the airways of the nose, [the] mouth, and pharynx," who together act as a flexible acoustic filter, responsible for '*articulation*,' or the demarcation of different phonemes and sounds.[101] Speech need not necessarily be phonated either, as in the case of whispered speech, where the source of acoustic energy is found subglottally by way of "frication noise created in a constriction, or a transient sound generated by rapid release of pressure when a vocal tract closure is opened."[102]

---

[100] Ville Pulkki and Matti Karjalainen. *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics* (Chichester, West Sussex: Wiley, 2015), 79.
[101] Lieberman and Blumstein, 4-5;
Pulkki and Karjalainen, 82-83;
The supralaryngeal vocal tract is also variously referred to in terms of its individual components, the pharynx, vocal tract and nasal tract.
[102] Pulkki and Karjalainen, *Communication Acoustics*, 90.

Speech can also be thought more abstractly, in terms of the widely-recognised schematic *source-filter* model put forward in Fant's *Acoustic Theory of Speech*.[103] In brief, Fant treats the above subglottal and laryngeal components as a singular 'source,' and the supralaryngeal vocal tract, functionally the "encoder of speech sounds," as a filter; as modelled by Pulkki, it also demarcates voiced and unvoiced sounds, as well as the acoustic action of both the vocal tract and lip radiation (see Figure 2.1).[104] This mode of sound production whereby certain frequencies are filtered out from a base signal can be considered analogous to *subtractive synthesis*.[105]

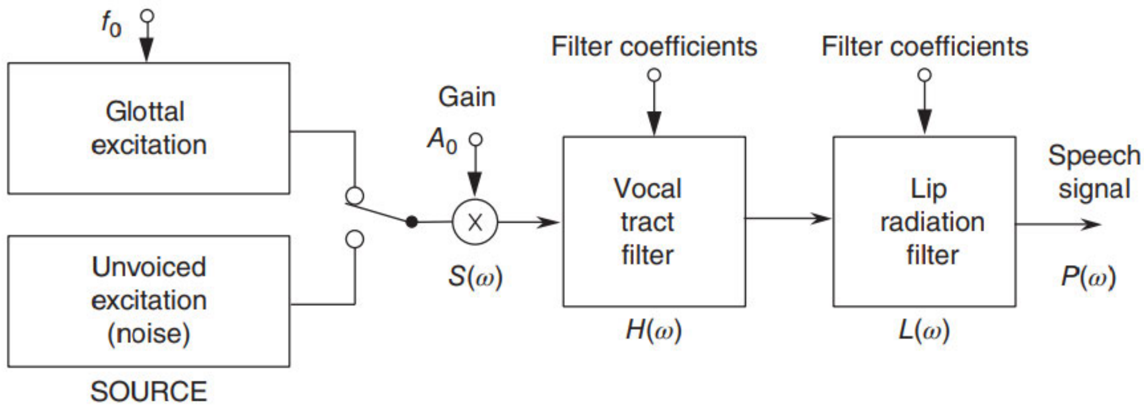


Figure 2.1. The *source-filter* model (or 'signal model') of speech production, as reproduced by Pulkki and Karjalainen in *Communication Acoustics*. Reproduced by permission of John Wiley & Sons, Inc.

Similar signal models have also been devised to describe the mechanism of acoustic musical instruments, albeit in mathematically complex technical detail, and primarily for the purpose of electronic synthesis.[106] Still, however, their concepts can be reduced and

[103] Gunnar Fant, *Speech Acoustics and Phonetics* (Dordrecht: Kluwer Academic Publishers, 2004), 15-21.
[104] Pulkki and Karjalainen, *Communication Acoustics*, 90-91.
[105] Perry R. Cook, "Sound Synthesis for Auditory Display," in *The Sonification Handbook*, ed. Thomas Hermann, Andy Hunt, and John G. Neuhoff (Berlin: Logos, 2011), 213-214.
[106] Marcelo Caetano and Xavier Rodet, "A Source-Filter Model for Musical Instrument Sound Transformation," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Kyoto, Japan: IEEE, 2012), 137-140;

Anssi Klapuri, "Analysis of Musical Instrument Sounds by Source-Filter-Decay Model," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Honolulu, HI, USA: IEEE, 2007), 1-4.

redeployed with respect to specific musical instruments, in a similar way to the previous example of speech (see Figure 2.2).[107]

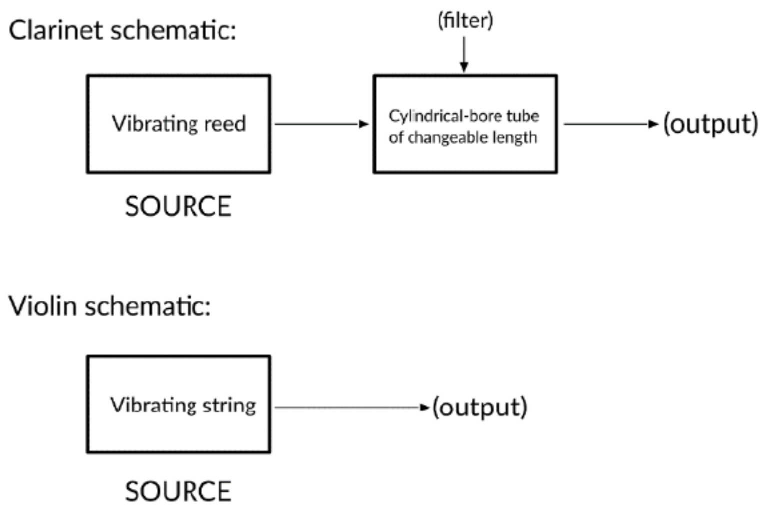


Figure 2.2. Highly simplified schematic model of a clarinet and violin in terms of (very pared back) subtractive synthesis.

The 'source' signal operates similarly amongst all three examples; filtration, by contrast, markedly does not. In the case of the violin, the material construction (and by extension, shape and size) of the resonating body is rigid and fixed, irrespective of what pitch is being played. In the case of a clarinet, the length of the resonating body can be finitely altered through the addition or venting of keys, but the instrument's material makeup, and, in a superficial sense, its shape, both largely remain the same; specific fingerings help a player produce specific pitches by invoking coincident resonant modes; different pitches produced this way still maintain an integral sense of timbral continuity, and timbre is for the most part not changeable independent from pitch or dynamic levels (at least within the physical instrument itself; this does not account for breathing, embouchure or other 'human factors').

[107] Note that these models are intentionally quite simplistic and do not capture all aspects of sound production – they aim mainly to capture the functional qualities of each instrument's main resonating body, for the sake of direct comparison to the human vocal tract; to give some examples of omitted factors, in the case of the violin, bow technique is largely ignored, and in the case of the clarinet, the effect of embouchure and mouth shape (or, the supralaryngeal vocal tract) on sound production is not considered.

By contrast, the human filter as exemplified in the shape of the mouth, position of the tongue, action or inaction of the lips *etc.* is highly flexible and dynamic, allowing for both a high degree of control and the potential for distinctive timbral discontinuity. Whereas most musical instruments, viewed as filters, can only be physically self-modified in terms of singular, quasi-linear paramaters (*e.g.* tube length for clarinet, head tension for the talking drum), speech production operates multiaxially, and is furthermore able to do so with complete independence from input and resultant pitch. In the context of effecting coherent speech, this is desirable; speech articulation is in fact made possible by the audible and characteristic timbral differences that exist between filter states, which are themselves realised as *phonemes*.

It is perhaps an obvious statement, but, owing to their rigid construction, musical instruments lack the innate capacity to capture the timbral diversity of speech, at least without augmentation via 'human factors,' or electronic intervention. Like-for-like recreation of the resonant mechanisms of speech within a fixed instrumental body, or – to put it in terms of synthrumentation – a *subtractive* model of speech resynthesis whereby basic instrumental tone is filtered for the purpose of timbral difference, is not innately possible within most existing instruments, and is all in all a methodological dead end. As anticipated by Barlow and others, human vocality cannot be instrumentally resynthesised by way of mechanistic imitation, and instead must be realised via *additive* processes.

2.2. Phonemes and formants

The techniques of additive resynthesis as found in synthrumentation employ a target material's identifiable sonic characteristics, usually by way of a spectrogram or similar, as a guide (for pitches and relative dynamic levels) for later imitative orchestration. In vocal synthrumentation, as already seen in Barlow's practice, it is useful to employ phonemes – the individual units of speech and language – as analytical targets, prior to any consideration of full sentences;[108] phonemes are constituted by both consonants and vowels, and vowels can in turn be categorised into monophthongs ("steady-state" with "no appreciable change in phonetic quality") and diphthongs (a monosyllabic vowel sound "including a change from one vowel quality to another").[109] (See Figure 2.3.)

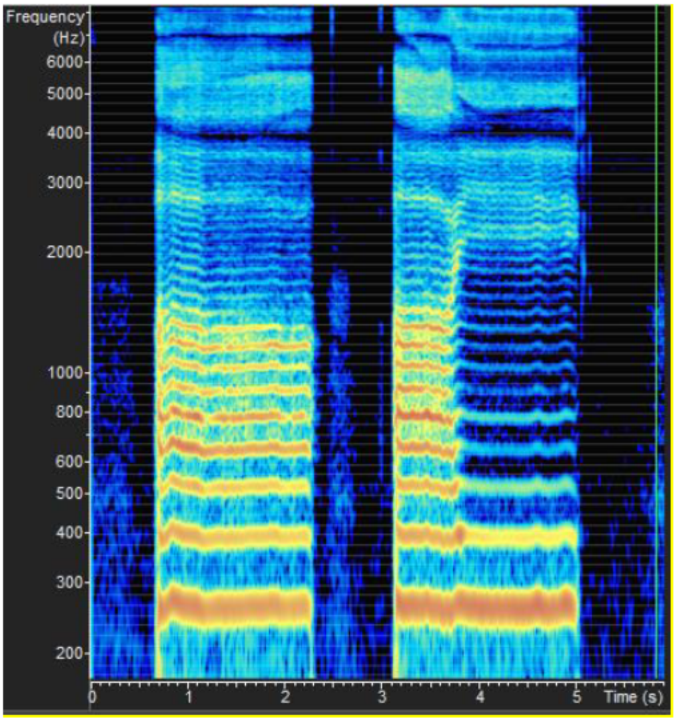


Figure 2.3. Two adjacent phonated vowel phonemes – the first is the [a] monophthong, and the second is the [aɪ] dipthong. Note how both are different sounds in spite of sharing nearly identical initial spectral states.

As previously mentioned, phonemic variation in part results from changes of resonance in the supralaryngeal tract; resonant bodies (including but not limited to that of humans), which acoustically will always (to varying degrees) allow through or filter out certain frequencies from a given input, accordingly possess deterministic frequency regions where sound energy is locally maximal. These frequency regions are called *formants*, and

[108] Barlow, "On the Spectral Analysis of Speech," 184-186.
[109] Peter Ladefoged and Sandra Ferrari Disner, *Vowels and Consonants* (Malden, MA: Wiley, 2012), 201-202; Ladefoged and Disner list "the [ɑ] in father" and "the [aɪ] in hide" as examples of monophthongs and diphthongs, respectively.

have been the focus and driver of many speech-acoustic studies to date, to the extent that “the controlled variation of formant frequencies is perhaps the single most important factor in human speech.”[110] It is common to label formant regions in a given speech sample sequentially (as F1, F2, F3, *etc.*) from lowest to highest frequency value (see Figure 2.4).

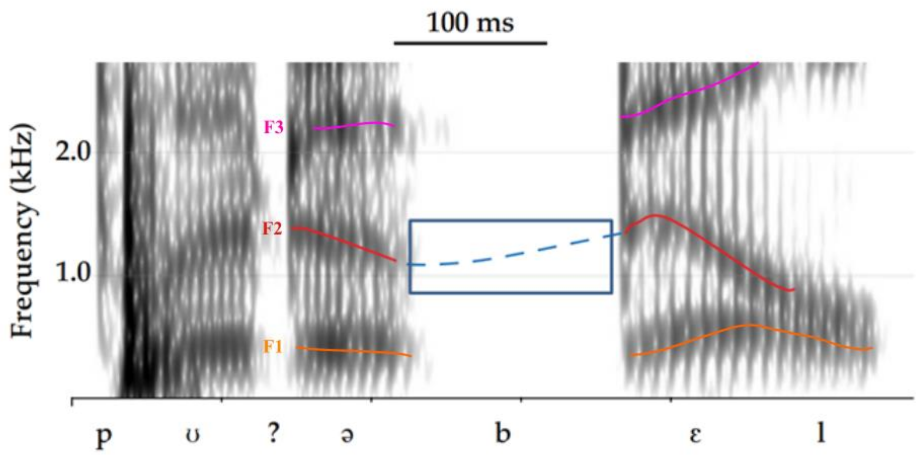


Figure 2.4. A standard use-case low resolution spectrogram, from Redford’s *Handbook of Speech Production*. F1, F2 and F3 are additionally annotated. Reproduced by permission of John Wiley and Sons, Inc.

Despite their “fundamental status” in speech acoustics, there are still technical aspects of formants that either remain elusive or are subject to approximation in common-practice.[111] For example, the formant frequencies of individual phonemes, whilst broadly similar, do differ slightly from person to person; several studies have gathered formant datasets across the spectrum of linguistic and sexual difference, but these are usually expressed in terms of a single numeric frequency value for each formant, and not above F4.[112] Likewise, the collection of data on formant bandwidths has not been a major focus; even the rare studies that do attempt to tabulate them do so non-specifically, with general reference to ‘vowels’

---

[110] Lieberman and Blumstein, 32;
Formants here are analogous with resonant modes as specifically used in 2.1.
[111] Marnix Van Soom and Bart de Boer, “A New Approach to the Formant Measuring Problem,” *Proceedings Vol. 33, No. 1* (2019): 1.
[112] James Hillenbrand, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler, “Acoustic Characteristics of American English Vowels,” *The Journal of the Acoustical Society of America, Vol. 95 No. 5* (1995): 3103;
David Deterding, “The Formants of Monophthong Vowels in Standard Southern British English Pronunciation,” *Journal of the International Phonetic Association Vol. 27, No. 1–2* (1997): 49-55.

rather than particular phonemic shapes (which often each entail their own individual sets of formant bandwidths).[113] Compounding all the above is the fact that formant measurement techniques are not fully standardised, and are still subject to a great degree of human choice and inherent error as a result.[114] As a consequence, the suitability of generalising formant research findings to other domains, including music, varies widely and needs to be considered from case to case.

Formants most prominently feature in vowels (both monophthongs and diphthongs, although they are also affective to varying degrees in consonants. This is particularly true of *approximants*, *e.g.* the sounds [w] and [y], which like vowels do not require the closure of the mouth, and like monophthongs are 'steady-state';[115] and is in a sense also true of *nasals*, who are conversely identifiable by the way they mask certain formant regions.[116] *Stops*, such as [d] and [b], express augmented formant regions when a closure (*i.e.* a stopping of airflow) "is being formed or is opening."[117] Finally, *fricatives e.g.* [f] and [s], bear some similarity to vowels by their capacity for 'steady-state' production, but the formant regions they possess (and their acoustic mechanisms) are completely unrelated to that of other phonemes, and are still the subject of ongoing clarification.[118]

---

[113] Raymond D. Kent and Houri K. Vorperian, "Static Measurements of Vowel Formant Frequencies and Bandwidths: A Review," *Journal of Communication Disorders* 74 (2018): 87;
[114] Van Soom and de Boer, "A New Approach to the Formant Measuring Problem," 1-2.
[115] Ladefoged and Disner, 53-54.
[116] *Ibid.*, 54-55
[117] *Ibid.*, 49-50
[118] Wiktor Jassem, "The Formants of Fricative Consonants," *Language and Speech Vol. 8 No. 1* (1965): 1–15;
Martine Toda, Shinji Maeda and Kiyoshi Honda, "Formant-Cavity Affiliation in Sibilant Fricatives," in *Interface Explorations: Turbulent Sounds: An Interdisciplinary Guide*, ed. Artemix Alexiadou and T. Alan Hall, (Berlin: Walter de Gruyter, 2010), 370-371.

2.3. Formant filtering and its effect on phonated and non-phonated (whispered) speech

Phonated vocal sounds are more or less periodic, [119] and therefore subject to the mode locking effect – their resultant spectrum tends to be concordant with the harmonic series as a result. Accordingly, phonated speech remains harmonic even when filtered through (inharmonic) formant regions (see Figure 2.5).

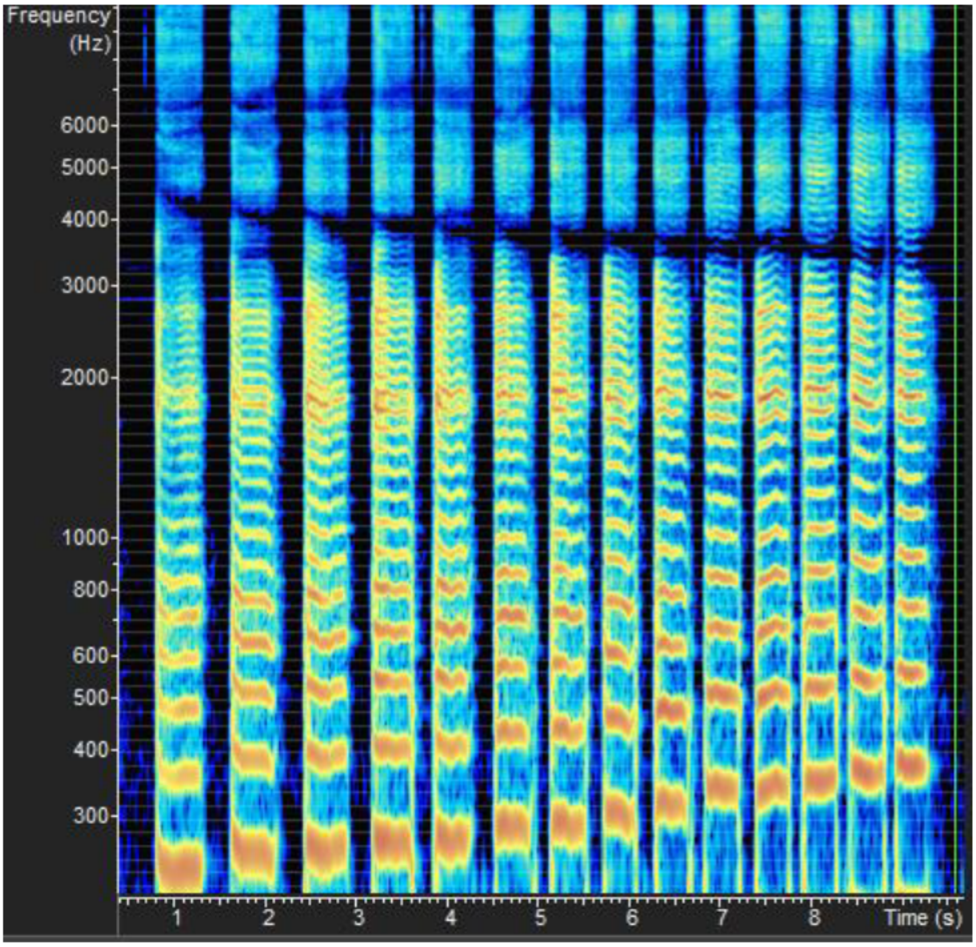


Figure 2.5. A phonated [ɛ] vowel, iterated across a series of rising pitches – note how the vocal spectra remain harmonic, even whilst being filtered through formant regions. It is also interesting to note how apparently static the formant regions are, in spite of the rising input pitch.

The phenomenon of identifiably harmonic phonated speech forms the basis of Barlow's working practice in *Im Januar* and *Orchideæ*. However, as previously discussed

[119] Pulkki and Karjalainen, *Communication Acoustics*, 80-82;
Paul Boersma, *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives* (The Hague: Holland Acad. Graphics, 1998), 16.

(see Ch. 1.4.), to view phonated speech (or any sound, for that matter,) as a solely harmonic construct is to downplay the importance of transients and residuals, and, as anticipated by Barlow, necessarily limits one from employing non-formant-expressing phonemes in synthrumental contexts.[120]

In Barlow's defence, a definitive capturing of phonated speech in synthrumentation is likely impossible by current standards – it is difficult to definitively separate harmonic and inharmonic content from a given, mono-dimensional speech signal, given both the aforementioned issues with strict formant identification as per Van Soom and de Boer (which in itself rules out any hope of meaningfully separating these components manually), as well as the current limitations of electronic methods.[121] So too, a cursory overview of said methods suggests that they also suffer from a similar 'human error' factor.[122]

By contrast, non-phonated, whispered speech is non-periodic, and, accordingly, entirely inharmonic; whispered speech's acoustic source (exhalation from the lungs) covers a broad-sweeping frequency spectrum with suppressed low-frequency regions, and in so doing both lacks a fundamental (F0) and also excites entire formant regions throughout their bandwidth, rather than just the discrete partials that coincidentally intersect them (see Figure 2.6).

[120] Barlow, "On the Spectral Analysis of Speech," 184.
[121] Every, "Separating Harmonic and Inharmonic Note Content," 5.
[122] Ibid., 1-4

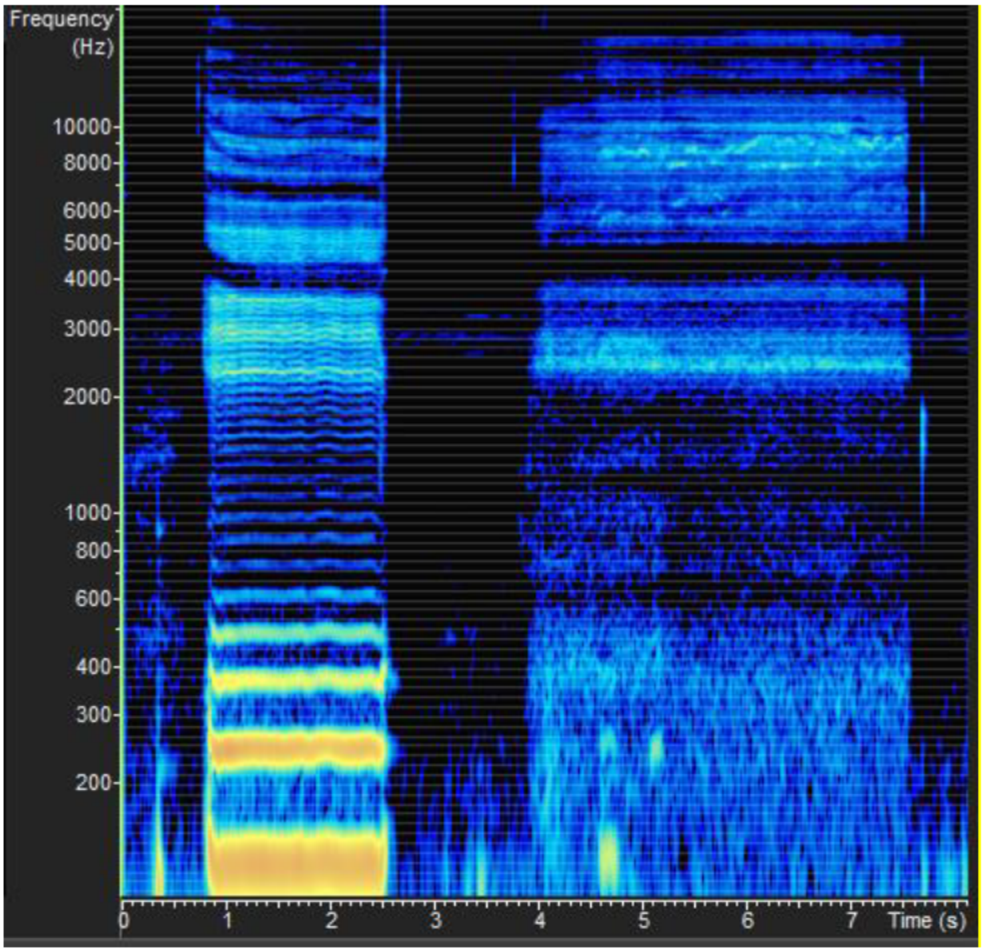


Figure 2.6. A phonated [i] vowel, followed by a non-phonated (whispered), identical [i] vowel (gain-adjusted). Note the general lack of distinctive peaks in the latter.

Here, sustained and residual components are subsumed into one, thus theoretically doing away with any need for their separate consideration during synthrumentative analytical and orchestrational processes. Note too that the formant regions of the phonated and non-phonated vowel differ significantly – F2 and F3 are both more compact in whispered [i], and F1 and F0 are barely present – even though they are ostensibly identical.

Similarly explicit differences between the formant structures (and spectral content) of whispered and phonated speech can be observed universally, in speakers across different cultures; strikingly, in spite of this, the two modes of communication still remain completely perceptually analogous,[123] and whispered speech maintains the capacity for conveying discrete tones and other intonational cues (including emotional cues), even though it lacks a fundamental (F0) frequency, and therefore, 'pitch'.[124] These qualities are still not fully understood; pertinently, although whispered formants are generally thought of as tending to be 'higher' than their phonated counterparts,[125] predictable or generalisable numerical relationships between in them in this regard (both in terms of peaks and bandwidths) have not been conclusively determined and are in fact the subject of contradictory findings.[126]

For all the above, whispered phonemes present as a promising initial candidate for a genuinely novel approach to vocal synthrumentation. They are relatively unexplored (both in a musical and linguistic sense), and also present a different set of considerations from Barlow's original method, particularly for their broad spectral bandwidths and lack of harmonic peaks; the synthrumentation of speech explicitly in terms of inharmonicity and formant bandwidths will accordingly be the aim of new methods devised in Chapter 3.

2.4. Perceptual considerations

Although they are a convenient target for resynthesis, phonemes (both harmonic and inharmonic, and their combined realisation in words and vocabulary) are not in and of

[123] Ken J. Kallail and Floyd W. Emanuel, "Formant-Frequency Differences Between Isolated Whispered and Phonated Vowel Samples Produced by Adult Female Subjects," *Journal of Speech, Language, and Hearing Research Vol. 27, No. 2* (1984): 245.
[124] Boon Pang Lim, "Computational Differences Between Whispered and Non-whispered Speech," PhD diss., University of Illinois at Urbana-Champaign, 2011, 35-37;
[125] Kallail and Emanuel, 245.
[126] Stanley J Wenndt, Edward J Cupples, and Richard M Floyd, "A Study on the Classification of Whispered and Normally Phonated Speech," *Proceedings of the 7th International Conference on Spoken Language Processing* (2002), 1;

Taisuke Ito, Kazuya Takeda, and Fumitada Itakura, "Analysis and Recognition of Whispered Speech," *Speech Communication Vol 45, No. 2* (2005): 141-143;

On a personal note: although it is outside the scope of this paper, I want to remark that the communicative and characteristic integrity of whispered speech as well as the simultaneous lack of empirical understanding surrounding it suggests fascinating aesthetic and philosophical implications when applied to musical contexts.

themselves the only important element of language: factors including grammatical structure, coarticulation, prosody, accent, intonation, and variability within, all play an important role in the structures and intelligibility of speech.[127] Cross-modal factors, such as visual cues, also have significant bearing on speech comprehension; Drager lists "lip-reading, facial expressions, and gestures" as a few examples.[128]

Barlow's exemplary vocal synthrumental works sidestep many of these, not only via a harmonic bias as discussed, but also through the composer's aim in the first instance to imprint speech onto pre-existing, arbitrary (non-speech-derived) melodies.[129] This paper, again, does not aim to make any empirical qualitative judgments on the efficacy of this or any other synthrumental methods in achieving their desired resynthesis, but, regarding *Im Januar*, Barlow himself has been somewhat critical in this regard: in past off-handedly remarking it "doesn't work that well," and later softening his assessment of the effect as "approximative."[130] It is not unreasonable to assume that these perceived shortcomings are a result not only of Barlow's orchestrational compromises, but also the non-incorporation of non-phonemic speech factors.

This naturally raises a number of questions with regards to the synthrumentation genre; which, as already established, intrinsically involves resynthesis, but *does not* aim to recreate target material with perfect hyper-accuracy. Instead, synthrumental practice prioritises only certain aspects of a target, both as a matter of practicality, as well as to better

---

[127] Jonathan Harrington and Steve Cassidy. *Techniques in Speech Acoustics*. *Vol. 8. Text, Speech and Language Technology* (Dordrecht: Springer Netherlands, 1999), 1-4;

[128] Dominique Simmons, Josh Dorsi, James W. Dias, and Lawrence D. Rosenblum. "Cross-Modal Transfer of Talker-Identity Learning." *Attention, Perception, & Psychophysics Vol. 83, No.* 1 (2021): 415–417;

Drager, Kathryn, and Joe Reichle, "Effects of Age and Divided Attention on Listeners' Comprehension of Synthesized Speech," *Augmentative and Alternative Communication Vol. 17, No. 2* (2001): 110.

[129] Kaske, "A Conversation with Clarence Barlow," 27.

[130] *Ibid.*;

Clarence Barlow in liner note to *Musica Algorithmica*, 6.

enable the conveyance of some kind of transcendental intent.[131] Importantly, however, in order for the latter outcome to come good, a synthrumentation still needs to be audibly, recognisably linked to its target source. (Otherwise the point of resynthesising a target source would be moot in the first instance.) The suggestion here is that, for a vocal synthrumentation to be 'successful', it needs to audibly resemble speech.

This is a relatively mundane statement, but it is in fact quite difficult to definitively resolve, especially when taking perceptual studies on speech into account. Encouragingly, base recognition that a sound 'constitutes speech' is a separate process from comprehension of meaning of speech,[132] implying that 'audible resemblance' can be achieved to an extent through reductive means, perhaps as in existing synthrumental practice. It is also understood that, neurologically speaking, automatic detection of speech occurs faster and more resource-efficiently than semantic interpretation, further supporting the efficacy of simplified means.[133]

However, research on the meaningful separation between these two categories (of speech detection and comprehension), let alone in the specific context of instrumental synthesis, is sparse: early experimental research found that subjects sometimes attributed vocal qualities to synthetic sounds that had "neither fundamental period nor formant structure, [and lacked] acoustic attributes [traditionally] assumed to underlie speech perception;" they also tended to do so more readily when prompted with the specific context "that they would hear a sentence;"[134] similar research has supported that even basic

---

[131] Barlow's *Im Januar* is, again, a good example of this, in that its reductive harmonic focus is intended to enable tonality within the piece; see Kaske, 25-27.
[132] O'Callaghan, Casey, "Against Hearing Meanings," *The Philosophical Quarterly Vol. 61, No.* 245 (2011): 801-807.
[133] Steven Greenberg and William A. Ainsworth, *Listening to Speech* (New York: Psychology Press, 2006) 413.
[134] Robert E. Remez et al., "Speech Perception Without Traditional Speech Cues," *Science, Vol. 212* (1981): 948-949.

adherence to formant regions is not a prerequisite for synthesised speech to come across.[135] However, more recent studies seem to emphasise the importance of learned contextualisation in enabling this to happen, and the role that mental ‘schema’ play in the recognition of speech objects.[136] In any case, it is impossible to know, by the standards of current research, the extent to which pre-contextualisation plays a role in determining the ‘resemblance’ of vocal synthrumentations;[137] by the same token, it is also unknowable which minimal selection of speech components need to be resynthesised or reorchestrated in context in order to achieve reflexive recognition of ‘speech’ in listeners.

Clearly, conclusive answers to these questions are far beyond the scope of this paper, and, in any case, an attempt to find them would require a readily available variety of vocal synthrumental techniques in the first instance. Given the open-ended nature of current scholarship, it is best here for this paper to continue to pursue whispered speech and resynthesis in terms of formant bandwidths, without overt consideration of perception (although this will be informally discussed in Ch. 3). So, too, as a matter of practicality and in order to avoid any discursive intersections with the above, I will be aiming to synthrument only phonemes in isolation.[138]

---

[135] C.J. Darwin and J.F. Culling, “Speech Perception Seen Through the Ear,” *Speech Communication Vol. 9, No. 5–6* (1990): 474.

[136] Radhika Aravamudhan, Andrew J. Lotto, and John W. Hawks, “Perceptual Context Effects of Speech and Nonspeech Sounds: The Role of Auditory Categories,” *The Journal of the Acoustical Society of America Vol. 124, No. 3* (2008): 1695–1703;

Simmons et al., 17-19;

Blauert, Jens, and Jonas Braasch, eds, *The Technology of Binaural Understanding* (Cham: Springer International Publishing, 2020), 51-52

[137] This is anecdotally echoed by Barlow himself, again in his interview with Kaske: “I have heard the computer synthesis [of *Im Januar*’s opening] many times; and recently while presenting a string ensemble performance of the piece I was suddenly able to hear speech at the beginning. It was a result of constant conditioning.”

[138] Again, this research is necessarily preliminary: more in-depth exploration of the potentials of vocal synthrumentation will be realised through further scholarly and compositional work, in due course.

CHAPTER 3.

A METHOD FOR SYNTHRUMENTING WHISPERED SPEECH

3.1. Clusters

Whispered speech occupies entire formant bandwidths, as opposed to just harmonic peaks. It follows that, in order to replicate the 'pitch' (or more appropriately, frequency) content of whispered speech in instrumental contexts (as opposed to phonated speech), different orchestrational strategies must be invoked from those used by Barlow and other synthrumentalists before, which tended to simply replicate harmonic peaks in instrumental contexts.

An intuitive solution to this would be to employ microtonal clusters within an orchestral group, so as to create a composite 'continuous bandwidth' of sound through instrumental means. The periodic partials of *ordinario* instrumental pitches, when spectrally analysed, superficially appear to have bandwidth distribution themselves (see Figure 3.1.) – if this is indeed reflective of reality, then theoretically a number of instruments playing 'overlapping' pitches could create the effect of a 'continuous bandwidth.'

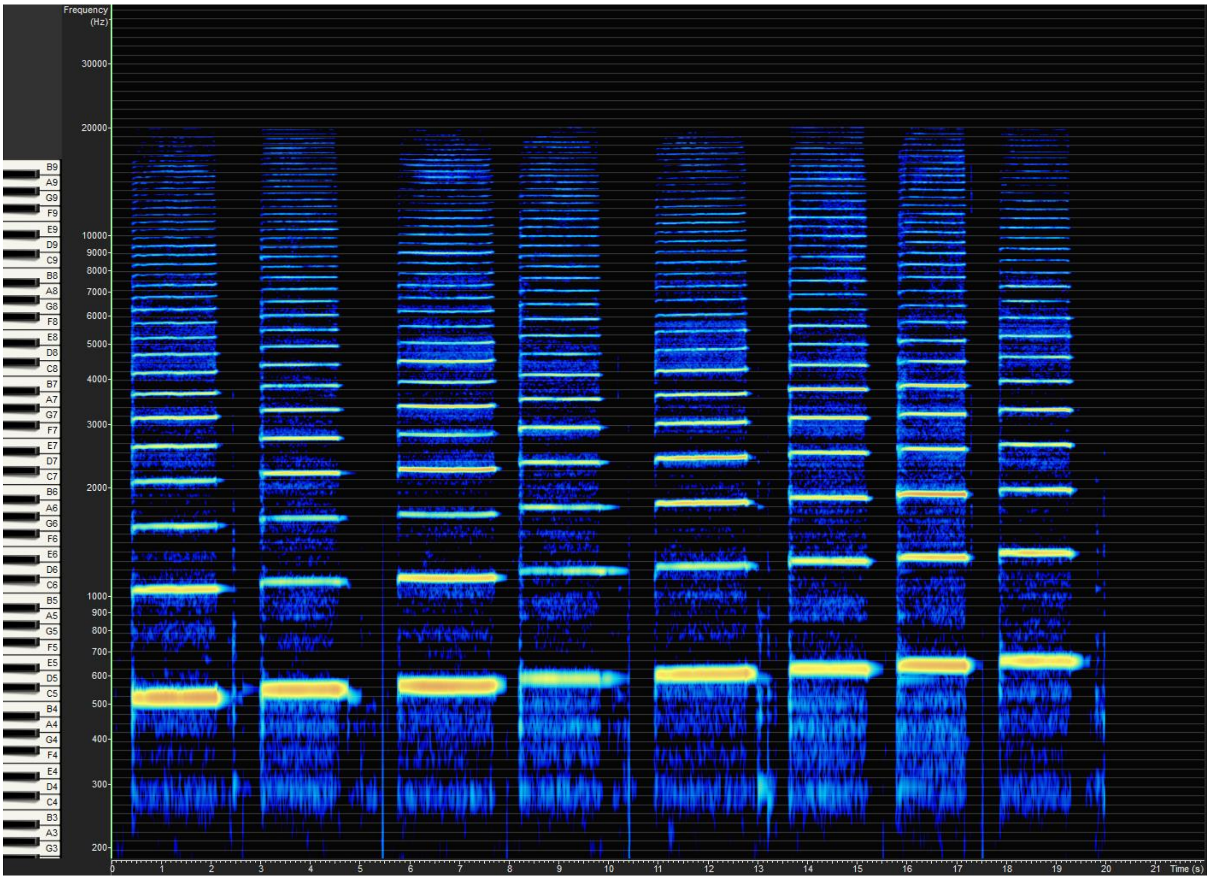


Figure 3.1 Spectrogram of a series of violin pitches, microtonally divided between C5 and E5 such that their fundamentals could 'overlap' – at least, visually speaking.

However, the apparent 'bandwidths' of individual harmonic partials is largely an artifact of the FFT process; conventional spectrograms, which are already subject to human choice and error (as per Ch. 2.3.), are also constrained by technical limitations; usage of 'high' and 'low' frequency resolutions does not entail the capture of more or less information, so much as different informational subsets.[139]

This is not necessarily fatal for a cluster-based method, however – instruments have access to non-periodic, broad-spectrum sounds themselves, and this will be further investigated in due course, in 3.2. At this juncture it is more pertinent instead to focus on the

[139] Harrington and Cassidy, 24-28.

practicalities and orchestrational considerations involved in the playing of clusters themselves.

Tone clusters are commonly thought of as an innovation of Henry Cowell, present in his piano works, where players were instructed to play a white-key or black-key subset of pitches (notationally bound by their outermost pitches) with their fist or forearm, primarily for “[registral] and percussive effect.”[140] There is of course an inherent limitation here, in that the intonation of a piano is fixed and its potential pitches finite; microtonal clusters were later implemented, with prescribed accidentals, in Scelsi’s music, and more recently have seen use in a more non-deterministic ‘upper-and-lower bound’ method similar to Cowell’s original in the music of Chaya Czernowin amongst many others.[141]

Dinescu has noted that the effectiveness (in terms of achieving compositional intent) of clusters as realised in live performance “depends on their semantic decoding by the performer.”[142] To take it further – and has been anticipated by prior synthrumentalists – an abundance of theoretical detail does *not* in and of itself directly translate into a parseable scoring method; simply demanding accuracy is not enough, and notation is itself another synthrumental aspect that needs to be mediated.

The required microtonal intervals between pitches in this new, ‘continuous’ method are unlikely to fall within convenient subsets of 12EDO, 24EDO, or others. Given both this, and the fact that microtonal intonation is to date still not taught as standard practice to

---

[140] Michael Hicks, “Cowell’s Clusters,” *The Musical Quarterly* Vol. 77, No. 3 (1993): 452.
[141] See Chaya Czernowin’s concerto for violoncello, *Guardian* (2017), for reference.
[142] Violeta Dinescu and Nelida Nedelcut, “The Cluster Effect – Connotations of Performance and Musical Notation,” *Proceedings of the International Symposium of Musicology “Pitfalls and Risks in Intra-Musical Communication: Terminology, Notation, Performance”* (2013): 28.

musicians, it would seem that a more non-deterministic approach to whispered phoneme synthrumentation is most appropriate.

Gould suggests a standardised notational practice for clusters, although she only speaks in terms of piano writing (see Figure 3.2).

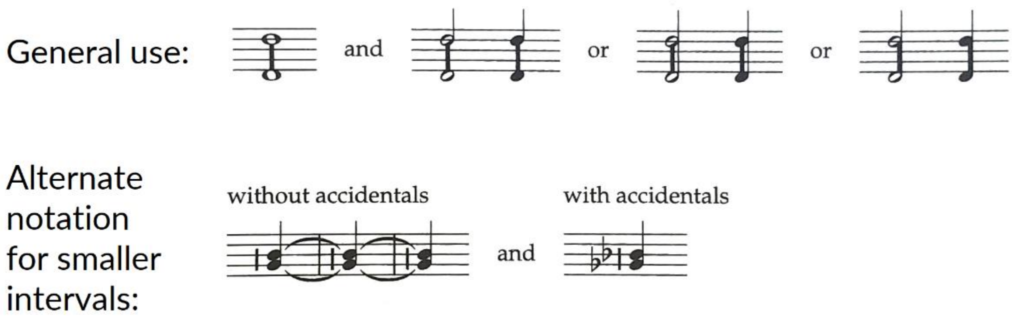


Figure 3.2. Elaine Gould's cluster notation from *Behind Bars*; additional annotations on left-hand side by author.

I propose that this form of notation be readapted for use within (microtonally capable) orchestral groups: a number of players in a given group, for example a violin section, would each be given the same part (*i.e.* just the one stave, instead of divisi) indicating a cluster between a lower and upper bound; in turn they would be asked to play a pitch within the bounds, proportional to their position within seating. For example, in a 6 player section, the first player on Desk 1 would play the upper bounding pitch, the second player on Desk 1 would play 'around 1/5th of the full interval lower,' and so on until the final player of Desk 3, who would play the lower bounding pitch. The result should be a relatively even spread of pitches, without the need to invoke a potentially impractical level of specificity. The efficacy of this notation can be observed in later implementations of the new synthrumental method.

3.2. Damping, or *rauschen*, as the chosen violin playing technique

In order for a clustral orchestration to successfully imitate continuous frequencies as found in whispered formants, it necessarily must employ non-periodic, microtonally accessible instrumental sounds, themselves with wide sonic bandwidths.

As discussed in Ch. 2.1., instrumental sound production is generally quite inflexible, particularly in inharmonicity-privileging contexts. For example, a timpani, whose resonance is inharmonic, has a very limited range (bound by head tension), and inflexible attack and decay characteristics. Wind and brass instruments can produce pure non-periodic sounds by having just air (and no other acoustic signal) blown through them – essentially being used as amplifier and filter of 'lung exhalation' sounds (as in Ch. 2.3.). Although there is greater potential here in terms of attack, sustain, and decay characteristics, the possible range of pitches is again limited by physical constraints (*i.e.* finitely extendable tube length, and also discrete keyed pitches). Furthermore, the use of an instrumental resonator as an 'amplifier' for the vocal cavity necessarily raises compatibility issues with the synthrumentation genre, which is wont to avoid the inclusion of an audio synthesis target itself.

I propose that this new whispered vocal synthrumentation method make use of bowed string instruments, using a particular technique of string *damping*, or '*rauschen*'.[143] This technique is available on any bowed string instrument; the player mutes the string by lightly placing multiple fingers in the left hand at the written pitch, at roughly harmonic pressure

[143] The latter term, *rauschen*, lit. "white noise," "hissing," or "rustling" (depending on context) has been used by the likes of Hans Abrahamsen – I personally prefer 'damping' for its connotation of the playing technique itself rather than resultant sound; this also avoids any potential confusion (particularly in German-language contexts) with other modes of production of white-noise-like sounds.

(see Figure 3.3.); the right hand bows *molto flautando* and dynamics are effected by changes in bow speed. The result is a wispy, white-noise-like sound, that has a small, detectable amount of pitch; spectrogram representations illustrate that a *damped* sound tends to express a wide frequency range around a central peak – much like the formant bandwidths they will be employed to imitate (see Figure 3.4).

Figure 3.3. Annotated photographs of *damped* playing action in the left hand, from front matter to *Rosy* (2020) for chamber ensemble, by author.

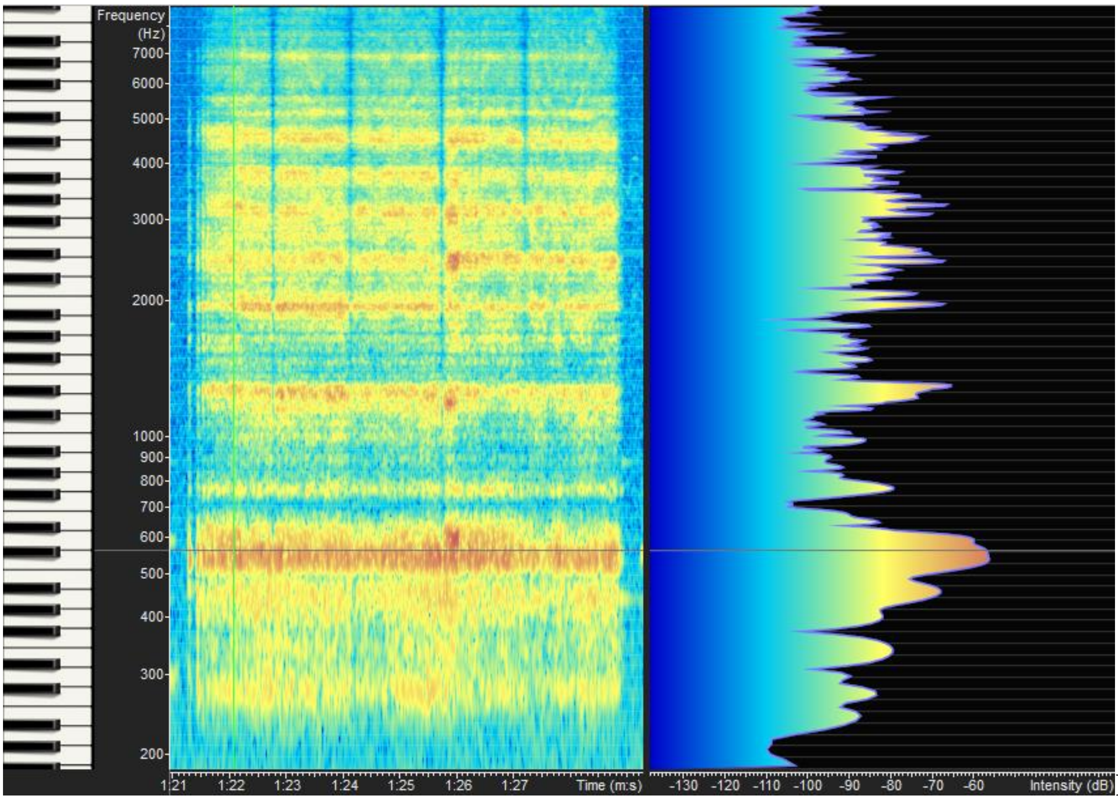


Figure 3.4. Spectrogram of a violin playing a C#5, damped. Midpoint shown to highlight bandwidth-like characteristics.

Use of string instruments in the cluster notation method is also convenient because of their immense range – a string can be bowed at a theoretically infinitely high pitch, which is particularly useful given the high frequency position of many vocal formants.[144] Likewise, bowed strings have very pliable attack and decay characteristics; played notes can be begun or cut off almost instantaneously in addition to gradual swells, providing flexibility in orchestration. The only major limitation of using *damping* to realise whispered formants is the need for flautando bow during playing; naturally a player will run out of bow quickly, especially when sustaining notes at (relatively) high dynamics. Bow changes can of course be staggered within a section, but this still requires care in the orchestrational process.

For the sake of illustration, here is a spectrogram of roughly the same violin pitch set presented *ordinario* in Fig. 3.1., instead played *damped* (see Figure 3.5.):

[144] Of course, the theoretical 'infinitely high' pitch of a bowed string is bound by physical contraints, including the width of the bow and resonant properties of the string itself.

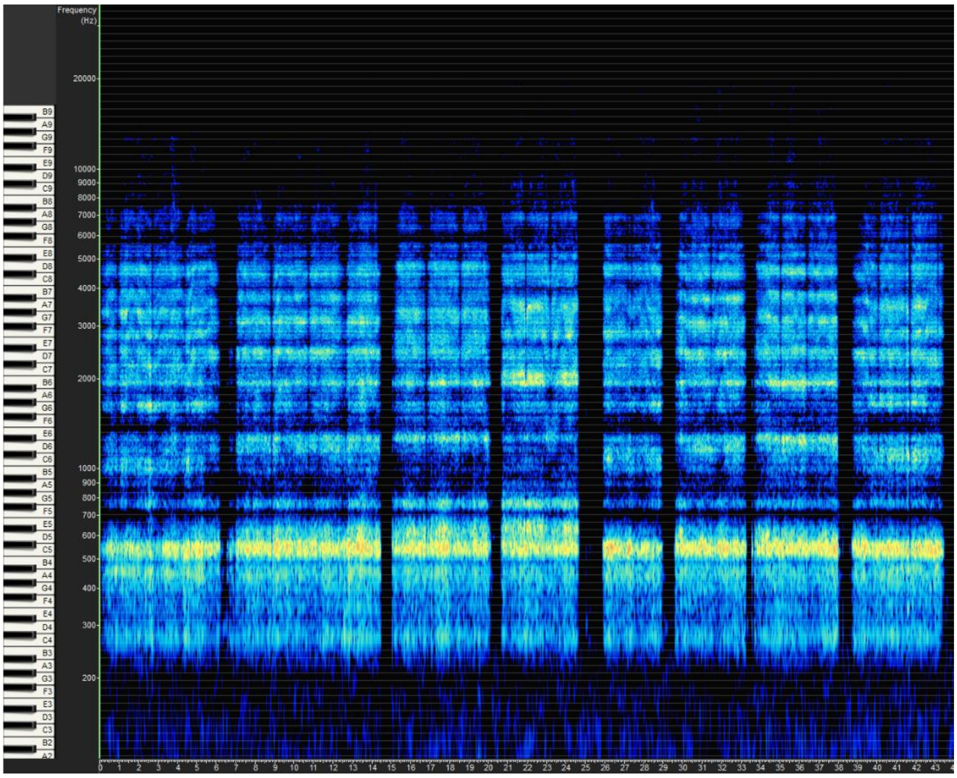


Figure 3.5. Spectrogram of a violin playing a C#5, damped. Midpoint shown to highlight bandwidth-like characteristics.

When the *ordinario* pitches of Fig. 3.1. are overlaid, in a similar fashion to how an 8-person violin section might interpret a notated cluster, individual peaks are both still audible and still visibly present in spectrogrammic representations; beating is visible, and a continuous frequency bandwidth is not achieved (see Figure 3.6).

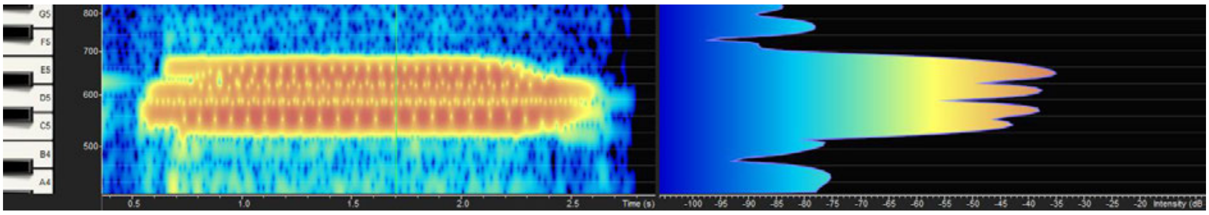


Figure 3.6. An 8-violin *ordinario* cluster from C_5 to E_5. Note the visible peaks and periodic interference (beating).

By contrast, the *damped* pitches of Fig. 3.5., when overlaid, produce a spectrogram that is in part visually similar to the 'continuous' formant bandwidths of whispered speech. The sound of the damped pitch cluster does not have any audible individual peaks; rather, the audible result is blended and unified (see Figure 3.7).

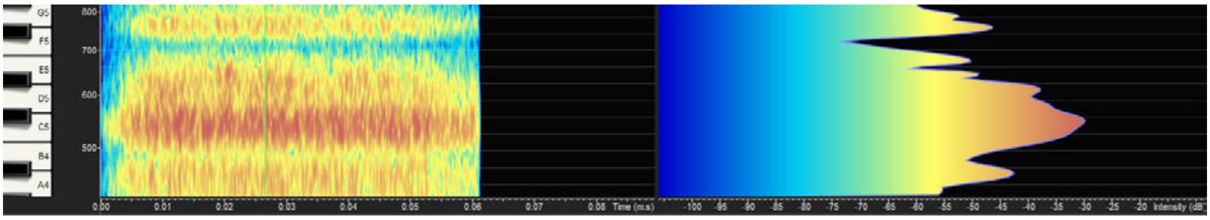


Figure 3.7. An 8-violin damped cluster from C_5 to E_5. Note the lack of visible peaks.

Further research is still required to determine how much bandwidth overlap is required between individual parts in order for the cluster to be effective, but as a basis for a new synthrumental practice the above results are more than adequate.

3.3. Phonemic choices, and limitations

Now, with a basic method formalised, all that is left to do is make a selection of what specifically it is that should be synthrumented. As established in Ch. 2.4., the goal at this point in time is to only resynthesise individual phonemes, not entire sentences.

At first, the ‘whispered’ method of synthrumentation seems more flexible than Barlow’s original, in that it is not harmonically bound and does not necessarily have to omit “phonemes containing noise spectra.”[145] This does not mean, however, that all phonemes can readily be synthrumented.

Firstly, stop consonants (such as [b] and [d]) are necessarily only perceivable with relation to an adjacent vowel sound; stops tend to momentarily shift the formants of a given vowel, in oft not entirely deterministic ways.[146] For the sake of simplicity, and because of their adjacency to speech proper, they cannot be included in this study.

[145] Kaske, 27.
[146] Ladefoged and Disner, 48-52.

Secondly, nasal consonants (such as [n] and [m]) are particularly difficult to express in terms of their whispered form; in natural whispered speech, they are by far the quietest consonants, and their perception is aided largely in part again by adjacency to vowels;[147] nasal phonemes also naturally suppress upper formant regions, which are crucial to the structures of whispered speech (which tends to not have a fundamental F0).[148] Because of the particularly heightened levels of uncertainty surrounding their transcription and implementation, they are also not considered here.

Chiefly left over are vowels (including both monophthongs and diphthongs) and a number of fricatives that can be produced 'steady state.' In Ch. 2.2 I discussed the general inconclusivity of existing phonated formant data sets, particularly for their tendency towards expression in terms of peaks; to further complicate matters, collections of data on the peak frequencies of whispered formants (which themselves are not the same as their phonated counterparts), let alone collections that include bandwidths, are nigh non-existent. As a result, synthrumental orchestrations herein necessarily must refer to self-generated, non-generalisable target audio.

Accordingly, a series of orchestrations for violins in multiples of 6-person groups (roughly analogous to a small section) was devised, employing the whispered formant bandwidth transcription and synthrumentation method detailed through this Chapter. An additional strategy was formed where multiple 6-person groups would be assigned to the same formant if it occurred in a low frequency region (so as to help with blending). The sonic target of these synthrumentations was a series of self-made recordings of various whispered phonemes. These orchestrations employed the *damping* technique, mediated by the

[147] Jovičić, Slobodan T., and Zoran Šarić, "Acoustic Analysis of Consonants in Whispered Speech," *Journal of Voice Vol. 22, No. 3* (2008): 268-273.
[148] Ladefoged and Disner, 54-55.

controlled-indeterminate microtonal cluster method also proposed. They can be found in Appendix A.

3.4. Recording method

In accordance with the aims of this research, these orchestrations were recorded in a live context. Due to COVID-19 room capacity restrictions at the time, a full string section could not be brought in to record – instead, to simulate the presence of more than 6 violinists, players were instructed to record multiple times per phoneme, in various fixed seating positions, whilst the microphone setup remained fixed. The result is that, collated in post, recorded material was aurally equivalent to a 'live' section of a larger number of players (save for the accumulation of background noise).

Two microphone setups were used to record: the first was a stereo pair of DPA 4060 miniature omni mics, attached to the ears of a Neumann KU 80 dummy head, with the intent of producing a quasi-binaural image. The second was a standard 20cm spaced pair of Schoeps MK5 cardioids. No microphone setup can like-for-like capture the exact aural experience of human ears, and thus the real, 'live' effect of the synthrumentations; in this case, the intent of having multiple setups was to provide a greater selection of recorded samples and to allow for more thorough assessment of sonic results after the fact.

Audio samples of the collated recorded output, in both microphone setupds, can be found as an attached media component.[149] Timestamps of specific phonemes are detailed in Appendix B. Not all available phoneme candidates were ultimately recorded due to time and

[149] A version of this paper with file embedded, for convenience, is available on the Royal College of Music Research Repository.

other practical constraints, but this provides an initial, fairly representative cross-section.

3.5. Informal discussion

Although this paper does not seek to objectively assess these synthrumentations, I want to make some personal, general observations here. Overall, I feel that the 'resynthesis' aspect of the orchestration worked well enough that a speech-like affect was innately perceptible. The question of specific phoneme accuracy is still left open – personally I detected a strong 'speech resemblance' in high-frequency phonemes such as [i] and [ɪ], and less so in phonemes with lower-frequency formants. Likewise, diphthong vowels were surprisingly effective, and did not sound particularly like wide-ranging glissandi, even though that was ostensibly what many of them were. Sibilants and fricatives were also quite communicative. However, my own personal perceptions are naturally impacted by my own prior knowledge of the technique, or 'constant exposure' to put it in Barlowian terms.

Likewise, differences between microphone setups were significant, although it is difficult at this juncture to identify exact, predictable trends in how this might be the case. All in all the sonic results of this research are highly compelling, if nothing else, and serve as a strong impetus for further specialised investigation, both in creative and academic contexts.

CONCLUSION

Overall, this study has been incredibly revealing on a number of fronts. A technical overview of *synthrumentation*, as a general term, has been provided, with several novel connections made between composers' varying practices both within and adjacent to the technique. Notably too is the consolidated understanding reached across them of the common role procedural simplification plays in their creative activity.

Likewise, although limited, the review and integration of scientific sources has allowed a clarification and in-depth understanding of the technical aspects of synthrumentation otherwise not immediately available or described in existing documentation. And, most importantly, this overall breadth of historical and scientific perspectives has been crucial in informing the creation of a novel method of synthrumentation, in all its intricacies.

As with any scholarly work, there have been inevitable limitations in the methods and findings of all the above. For one, due to an initial impetus from Barlow (an Anglophone composer with working practice in Germany) much historical and cultural discussion has centred on Western musical practices, and there is no doubt much that has been unintentionally omitted from the historical overview in Ch. 1, let alone other representative works within, prior, or adjacent to the synthrumentation 'genre'.

As already acknowledged in Ch. 2, the formalised integration of scientific sources with musical practice in a scholarly context is a far-reaching, specialised task, and one beyond the scope of this study. However, the review within this chapter does establish the relevance of certain existing findings in recent research, and certainly further suggests other areas not-yet-well explored that might now warrant more attention (*e.g.* whispered formant tables, the demarcations of linguistic and pre-linguistic speech recognition, and so on).

In many ways, this research paper serves as a preliminary study – there are still many aspects of 'synthrumentation' that need to be further interrogated, particularly in terms of the technique's musical and aesthetic implications. There are no doubt refinements required of the method presented, too; presentation of the phonemic inventory of human speech, let alone just the English language, was incomplete; functional use (both musical and semantic) of these synthrumentations was not demonstrated; and there are likely countless other factors to consider.

However, in any case, the definitive outcome of this paper has at its core been the proposal and implementation of a new, affectively compelling sonic material, with potential myriad uses in actual musical contexts. The practices of synthrumentation, which have oft been underappreciated since the turn of the millennium, ought to be the subject of future scholarly and compositional work, and I hope that this paper and its presented methods can be a strong impetus and source of encouragement in that regard.

BIBLIOGRAPHY

Abbate, Carolyn. "Debussy's Phantom Sounds." *Cambridge Opera Journal Vol. 10 No. 1* (1998): 67-96.

Ablinger, Peter. "Phonorealism." Accessed June 1, 2021. https://ablinger.mur.at/phonorealism.html.

Anderson, Julian. "A Provisional History of Spectral Music." *Contemporary Music Review 19, No. 2* (2000): 7–22.

Aravamudhan, Radhika, Andrew J. Lotto, and John W. Hawks. "Perceptual Context Effects of Speech and Nonspeech Sounds: The Role of Auditory Categories." *The Journal of the Acoustical Society of America Vol. 124, No. 3* (2008): 1695–1703.

Barlow, Clarence. *Musica Algorithmica*. Ensemble Köln, Robert H.P. Platz, Iceland Symphony Orchestra, Hermann Bäumer, Südwestfunkorchester Baden-Baden, Ingo Metzmacher, Ensemble Modelo62, and Ezequiel Menalled. Maria de Alvear World Edition 0034, 2018. 2 CDs.

———————. "On Music Derived from Language." *International Journal of the International Institute for Advanced Studies in Systems Research and Cybernetics, Vol. 9 No. 1* (2009): 29-33. http://clarlow.org/wp-content/uploads/2016/10/On-Music-Derived-from-Language.pdf.

———————. "On the Spectral Analysis of Speech for Subsequent Resynthesis by Acoustic Instruments." *Forum Phoneticum 66: Festschrift Georg Heike* (1998): 183-190.

Barrett, G. Douglas. *Window Piece: Seeing and Hearing the Music of Peter Ablinger*. Self-published.

Blauert, Jens, and Jonas Braasch, eds. *The Technology of Binaural Understanding*. Cham: Springer International Publishing, 2020.

Boersma, Paul. *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives*. The Hague: Holland Acad. Graphics, 1998.

Braden, Alistair C. P., Michael J. Newton, and D. Murray Campbell. "Trombone Bore Optimization Based on Input Impedance Targets." *The Journal of the Acoustical Society of America Vol. 125, No. 4* (2009): 2404–12.

Caetano, Marcelo, and Xavier Rodet. "A Source-Filter Model for Musical Instrument Sound Transformation." In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 137–40. Kyoto, Japan: IEEE, 2012.

Carpentier, Grégoire, and Jean Bresson. "Interacting with Symbol, Sound, and Feature Spaces in Orchidée, a Computer-Aided Orchestration Environment." *Computer Music Journal Vol. 34, No. 1* (2010): 10–27.

Chaurasiya, Himanshu. 'Time-Frequency Representations: Spectrogram, Cochleogram and Correlogram'. *Procedia Computer Science Vol. 167* (2020): 1901–10.

Chen, Andrew. *Rosy*. Unpublished, 2020.

Cook, Perry R. “Sound Synthesis for Auditory Display.” In *The Sonification Handbook*, ed. Thomas Hermann, Andy Hunt, and John G. Neuhoff, 197-235. Berlin: Logos, 2011.

Darwin, C.J., and J.F. Culling. “Speech Perception Seen Through the Ear.” *Speech Communication Vol. 9, No. 5–6* (1990): 469-75.

Deterding, David. “The Formants of Monophthong Vowels in Standard Southern British English Pronunciation.” *Journal of the International Phonetic Association Vol. 27, No. 1–2* (1997): 47–55.

Dinescu, Violeta, and Nelida Nedelcut. “The Cluster Effect – Connotations of Performance and Musical Notation.” *Proceedings of the International Symposium of Musicology “Pitfalls and Risks in Intra-Musical Communication: Terminology, Notation, Performance.”* (2013): 17-29.

Drager, Kathryn, and Joe Reichle. “Effects of Age and Divided Attention on Listeners’ Comprehension of Synthesized Speech.” *Augmentative and Alternative Communication Vol. 17, No. 2* (2001): 109–19.

Dufourt, Hugues. “Musique spectrale.” In *Musique, pouvoir, e´criture*, 289–294. Paris: Christian Bourgois, 1991.

Every, Mark. “Separating Harmonic and Inharmonic Note Content From Real Mono Recordings.” *University of York, Department of Electronics* (2005): 1-5. https://www-users.york.ac.uk/~jes1/EveryDMRN05.pdf.

Fant, Gunnar. *Speech Acoustics and Phonetics*. Dordrecht: Kluwer Academic Publishers, 2004.

Féron, François-Xavier. “Gérard Grisey: première section de *Partiels* (1975).” *Genesis, Vol. 31* (2010): 77-97.

———————. “The Emergence of Spectra in Gérard Grisey's Compositional Process: From *Dérives* (1973–74) to *Les espaces acoustiques* (1974–85).” *Contemporary Music Review, Vol. 30 No. 5* (2011): 343-375.

Fillerup, Jessie. “Purloined Poetics: The Grotesque in the Music of Maurice Ravel.” PhD diss. University of Kansas, 2009.

Fineberg, Joshua. “Musical Examples.” *Contemporary Music Review Vol. 19, No. 2* (2000): 115–34.

Fletcher, N. H. “Mode Locking in Nonlinearly Excited Inharmonic Musical Oscillators”. *The Journal of the Acoustical Society of America Vol. 64, No. 6* (1978): 1566-69.

———————. “The Nonlinear Physics of Musical Instruments.” *Reports on Progress in Physics Vol. 62, No. 5* (1999): 723-64.

Gould, Elaine. *Behind Bars*. London: Faber, 2011.

Greenberg, Steven and William A. Ainsworth. *Listening to Speech.* New York: Psychology Press, 2006.

Grisey, Gérard. *Les Espaces Acoustiques*. Garth Knox, Asko|Schönberg, WDR Sinfonieorchester Köln, and Stefan Asbury. KAIROS 0012422KAI, 2005. 2 CDs.

———————. *Partiels*. Milan: Ricordi, 1976.

Harrington, Jonathan, and Steve Cassidy. *Techniques in Speech Acoustics*. *Vol. 8. Text, Speech and Language Technology*. Dordrecht: Springer Netherlands, 1999.

Harrison, Peter M. C., and Marcus T. Pearce. "Simultaneous Consonance in Music Perception and Composition." *Psychological Review Vol. 127, No. 2* (2020): 216-44.

Harvey, Jonathan. *Speakings*. London: Faber, 2008.

Herd, Judith Ann. "The Neonationalist Movement: Origins of Japanese Contemporary Music." *Perspectives of New Music Vol. 27, No. 2* (1989): 118-163.

Henrot, Jérémie. "Speakings – Jonathan Harvey: Technical Rider for Performances and Rehearsals." Accessed May 23, 2021. https://brahms.ircam.fr/media/uploads/Tech_Rider_-_Generic_for_Speakings.pdf.

Hicks, Michael. "Cowell's Clusters." *The Musical Quarterly* Vol. 77, No. 3 (1993): 428–58.

Hillenbrand, James, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler. "Acoustic Characteristics of American English Vowels." *The Journal of the Acoustical Society of America, Vol. 95 No. 5* (1995): 3099-111.

Ihde, Don. *Listening and Voice: Phenomenologies of Sound, 2nd Ed.* Albany: SUNY Press, 2007.

Ito, Taisuke, Kazuya Takeda, and Fumitada Itakura. "Analysis and Recognition of Whispered Speech'. *Speech Communication Vol 45, No. 2* (2005): 139–52.

Jassem, Wiktor. "The Formants of Fricative Consonants." *Language and Speech Vol. 8, No. 1* (1965): 1–16.

Jovičić, Slobodan T., and Zoran Šarić. "Acoustic Analysis of Consonants in Whispered Speech." *Journal of Voice Vol. 22, No. 3* (2008): 263–74.

Kallail, Ken J., and Floyd W. Emanuel. "Formant-Frequency Differences Between Isolated Whispered and Phonated Vowel Samples Produced by Adult Female Subjects." *Journal of Speech, Language, and Hearing Research Vol. 27, No. 2* (1984): 245–51.

Kaske, Stephan. "A Conversation with Clarence Barlow." *Computer Music Journal Vol. 9, No. 1* (1985): 19–28.

Kent, Raymond D., and Houri K. Vorperian. "Static Measurements of Vowel Formant Frequencies and Bandwidths: A Review." *Journal of Communication Disorders* 74 (2018): 74–97.

Klapuri, Anssi. "Analysis of Musical Instrument Sounds by Source-Filter-Decay Model." In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Honolulu, HI, USA: IEEE, 2007.

Koenig, W. et al. “The Sound Spectrograph.” *The Journal of the Acoustical Society of America, Vol. 18 No. 1.* (1946): 19-49.

Krier, Yves. “Partiels, de Gérard Grisey, Manifestation D'une Nouvelle Esthétique.” *Musurgia Vol. 7 No. 3-4* (2000): 145-172.

Ladefoged, Peter, and Sandra Ferrari Disner. *Vowels and Consonants*. Malden, MA: Wiley, 2012.

Lieberman, Philip, and Sheila Blumstein. *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge: Cambridge University Press, 1988.

Lim, Boon Pang. “Computational Differences Between Whispered and Non-Whispered Speech.” PhD diss. University of Illinois at Urbana-Champaign, 2011.

McAdams, Stephen, and Bruno L. Giordano. *The Perception of Musical Timbre*. Edited by Susan Hallam, Ian Cross, and Michael Thaut. Oxford: Oxford University Press, 2012.

Moscovich, Viviana. “French Spectral Music: An Introduction.” *Tempo, New Series, No. 200* (1997): 21-27.

Murail, Tristan. “Gondwana.” *Tristan Murail*. Accessed May 16, 2021. https://www.tristanmurail.com/en/oeuvre-fiche.php?cotage=TR1572.

———————. “Spectra and Pixies.” *Contemporary Music Review Vol 1, No. 1* (1984): 157–70.

———————. *Tristan Murail: Serendib / L'esprit des dunes / Désintégrations*. Ensemble Intercontemporain, David Robertson. Accord AC4653052, 1996. 1 CD.

Nouno, Gilbert, Arshia Cont, Grégoire Carpentier, and Jonathan Harvey. “Making an Orchestra Speak.” *Sound and Music Computing* (2009): 1-6. https://hal.inria.fr/hal-00839067/document.

O’Callaghan, Casey. “Against Hearing Meanings.” *The Philosophical Quarterly Vol. 61, No.* 245 (2011): 783–807.

Poller, Tom Rojo. “Clarence Barlow’s Technique of ‘Synthrumentation’ and Its Use in *Im Januar Am Nil*.” *Tempo Vol. 69 No. 271* (2015): 7-23.

Pulkki, Ville and Matti Karjalainen. *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics.* Chichester, West Sussex: Wiley, 2015.

Puri, Michael J. *Ravel the Decadent: Memory, Sublimation, and Desire*. Oxford: Oxford University Press, 2011.

Ravel, Maurice. *Daphnis et Chloé*. Paris: Durand, 1913.

Redford, Melissa A. *The Handbook of Speech Production*. Chichester, West Sussex: Wiley, 2015.

Remez, Robert E., Philip E. Rubin, David B. Pisoni and Thomas D. Carrell. “Speech Perception Without Traditional Speech Cues.” *Science, Vol. 212* (1981): 947-950.

Ritsch, Winfried, Peter Ablinger, and Thomas Musil. “Robotic Piano Player Making Pianos Talk.” *Institute for Electronic Music and Acoustics Graz* (2011).

Rodrigues, Felipe Galhardi. “Representações do Caos na Música do Século XVIII,” in *Linguística, Letras e Artes: Cânones, Ideias e Lugares 2*, edited by Ivan Vale De Sousa, 119–27. Ponta Grossa, Brazil: Atena Editora, 2020.

Rose, François. “Introduction to the Pitch Organization of French Spectral Music.” *Perspectives of New Music Vol. 34, No. 2* (1996): 6-39.

Sachs, Curt. *The Rise of Music in the Ancient World, East and West*. New York: Dover, 2008.

Shimizu, Yoshihiko. “The Creative Quest into Temple Bell Sonorities: Works of *Musique Concrète* by Toshiro Mayuzumi.” *Contemporary Music Review Vol. 37, No. 1* (2018): 36–48. https://doi.org/10.1080/07494467.2018.1453335

———————. “Serial Technique in Toshiro Mayuzumi’s ’Campanology Effect’,” *Journal of the Musicological Society of Japan, Vol. 56, No. 1* (2010): 26-37.

Simmons, Dominique, Josh Dorsi, James W. Dias, and Lawrence D. Rosenblum. “Cross-Modal Transfer of Talker-Identity Learning.” *Attention, Perception, & Psychophysics Vol. 83, No.* 1 (2021): 415–34.

Smirnov, Andrey and Liubo Pchelkin. *Russian Pioneers of Sound Art in the 1920s.* Catalogue of the exhibition “Red Cavalry: Creation and Power in Soviet Russia between 1917 and 1945.” Madrid: La Casa Encendida, 2011.

Takakura, Yuriko. “A Comparison of the Compositional Process between the *Nirvana Symphony* and the *Mandala Symphony*: An Analysis of the “Campanology Documents.”” *Ongakugaku, Vol. 63, No. 2* (2017): 61-77.

———————. “A Visual Analysis Methodology for Music Compositional Process with Sound Resynthesis.” PhD diss. Keio University, 2019.

Thaut, Michael. “Music As Therapy in Early History.” In *Music, Neurology and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies*. Edited by Eckart Altenmüller, Stanley Finger and François Boller, 143-158. Amsterdam: Elsevier, 2015.

Toda, Martine, Shinji Maeda and Kiyoshi Honda. “Formant-Cavity Affiliation in Sibilant Fricatives.” In *Interface Explorations: Turbulent Sounds: An Interdisciplinary Guide*, ed. Artemix Alexiadou and T. Alan Hall, 343-374. Berlin: Walter de Gruyter, 2010.

Van Soom, Marnix, and Bart de Boer. ‘A New Approach to the Formant Measuring Problem’. *Proceedings Vol. 33, No. 1* (2019): 29. https://doi.org/10.3390/proceedings2019033029.

Verma, Tony S., and Teresa H. Y. Meng. “Extending Spectral Modeling Synthesis with Transient Modeling Synthesis.” *Computer Music Journal Vol. 24, No. 2* (2000): 47–59.

Walker, D. P. *Spiritual and Demonic Magic: From Ficino to Campanella*. Pennsylvania: PSU Press, 2003.

Watkins, Holly. *Musical Vitalities: Ventures in a Biotic Aesthetics of Music*. Chicago: The University of Chicago Press, 2018.

Wenndt, Stanley J, Edward J Cupples, and Richard M Floyd. "A Study on the Classification of Whispered and Normally Phonated Speech." *Proceedings of the 7th International Conference on Spoken Language Processing* (2002).

Wolfe, Joe. "How Harmonic are Harmonics?" *University of New South Wales School of Physics*. Accessed May 21, 2021. https://newt.phys.unsw.edu.au/jw/harmonics.html.

APPENDIX A: Synthrumented whispered phonemes, orchestrated for violins

The present orchestrat ons attempt to capture the sounds of wh spered speech phonemes, v a synthrumenta processes A number of se f generated record ngs of wh spered speech were manua y transcr bed, and readapted to use the c uster notat on a d out n Chapter 3 Each ne shou d be p ayed by a sect on of 6 or more v o ns; occas ona y a th rd sect on (n pos t on C) s requ red to smooth out ow formants

Regard ng c usters Ind v dua p ayers shou d nd scr m nate y p ay a p tch w th n the bounded c uster, rough y proport ona to the r seat ng pos t on (*e.g.* the r ght most p ayer cou d p ay the ower bound, the eft most the ower bound, and then proport ona y p ayers fi n the p tches n between, w thout any med at on from tun ng dev ces)

A though the exact m crotona p tches w th n c usters aren t prescr bed, m crotona acc denta s are st used for the upper and ower bounds of a c uster They nc ude standard quartertones, as we as s xth tones, represented by an up or down arrow a terat on

P ayers shou d p ay *damped* throughout a wr tten passages N B a so s ght y non standard use of 8va and 15ma nd cat ons (on y used n study score and not parts)

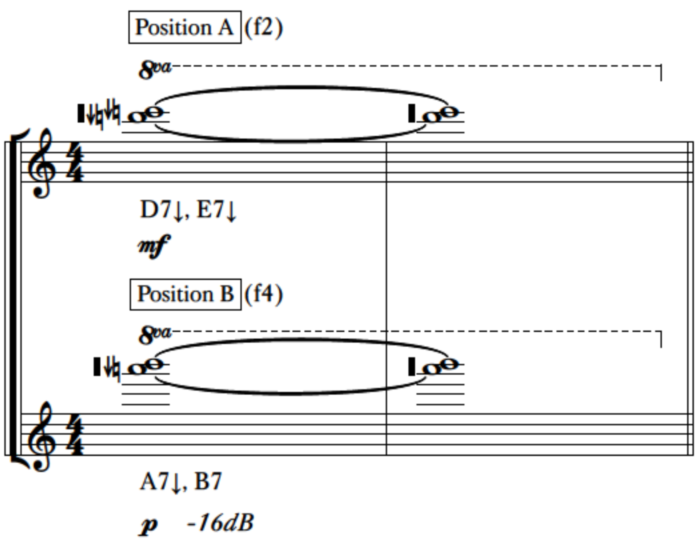
Orchestrat ons have add t ona y been annotated w th nformat on about re at ve (e orchestrated) and abso ute (e data sourced) dynam c eve s, as we as rough est mat ons of the formants be ng captured n each stave These are presented for the sake of the research paper, and norma y on y mus ca dynam cs wou d be nc uded n p ayers parts

L kew se, pos t ons A, B, and C are re evant on y to the mu t track ng method be ng used n the record ng sess on tse f (due to COVID 19 capac ty restr ct ons at the t me)

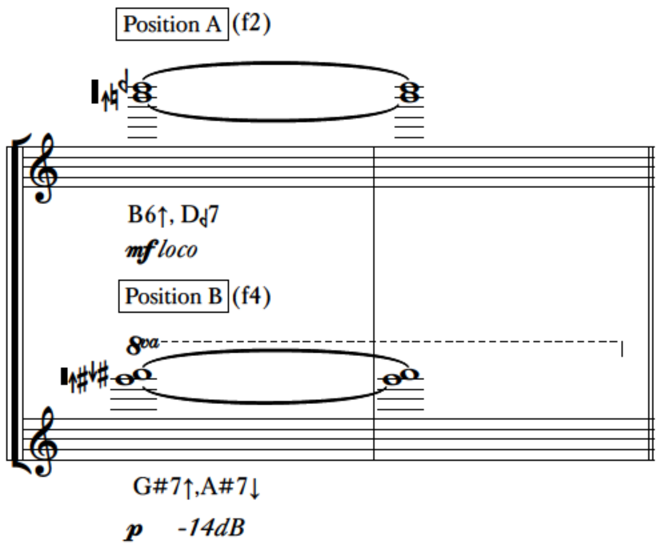
♩= 60

1. MONOPHTHONGS (1/3)

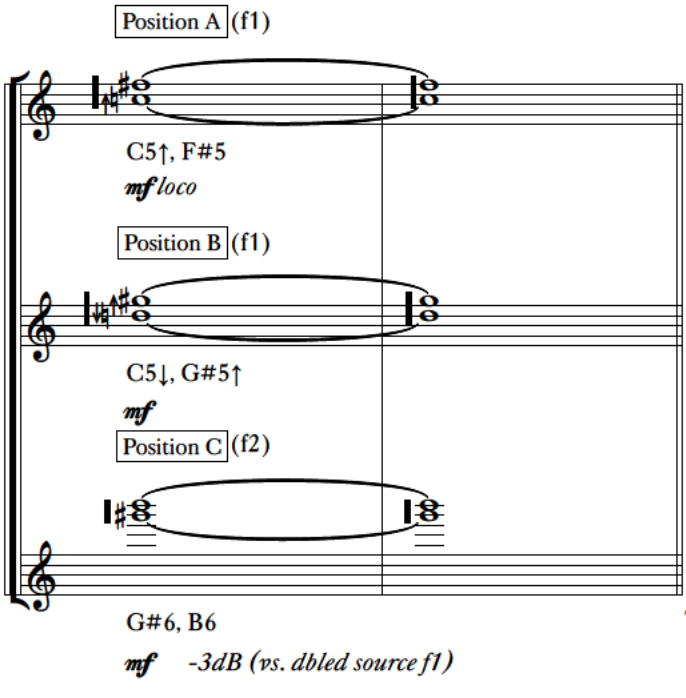
1.1 Vowel: [i] as in "need"



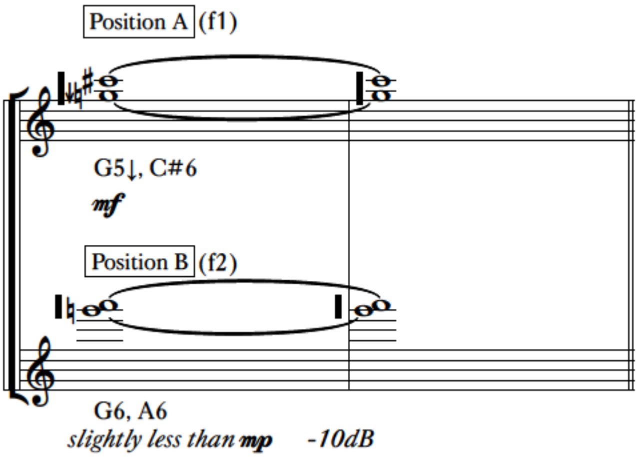
1.2 Vowel: [ɪ] as in "bid"



1.3 Vowel: [ɛ] as in "bed"



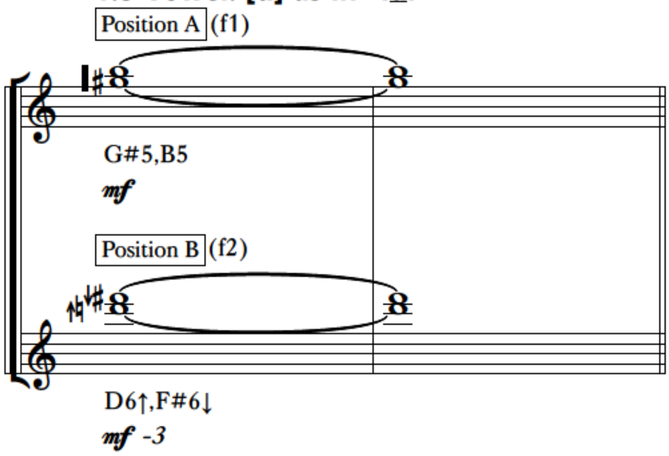
1.4 Vowel: [æ] as in "back"



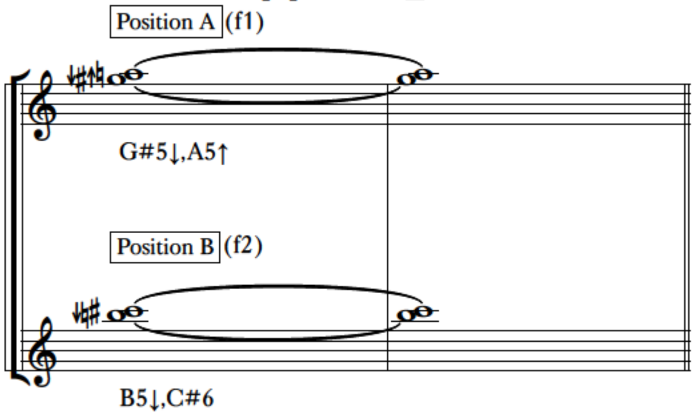
♩= 60

1. MONOPHTHONGS (2/3)

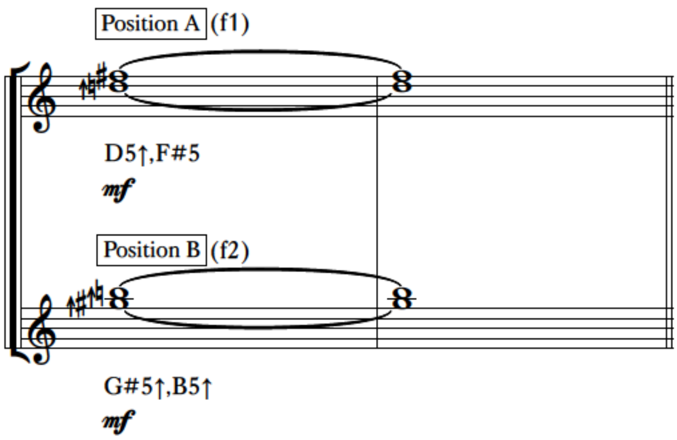
1.5 Vowel: [ɑ] as in "far"



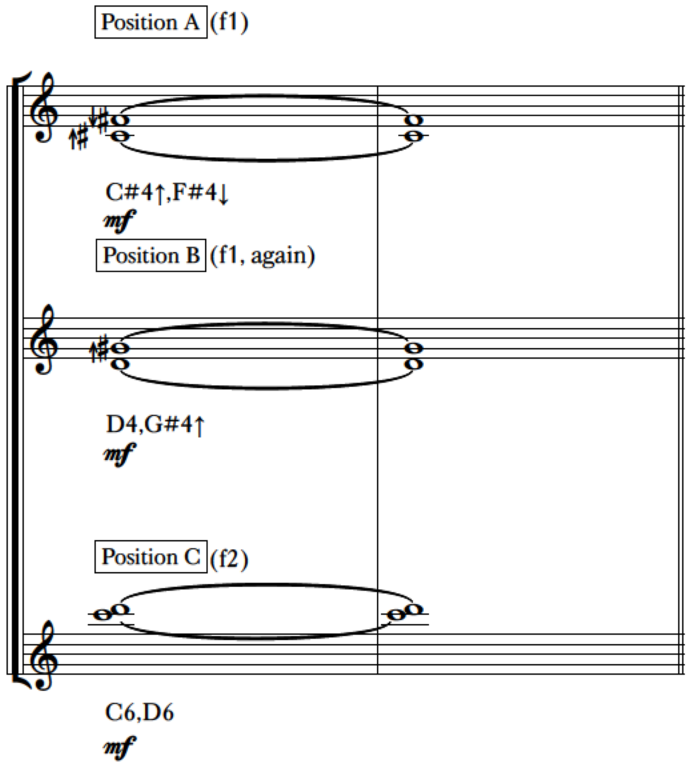
1.6 Vowel: [ɒ] as in "socks"



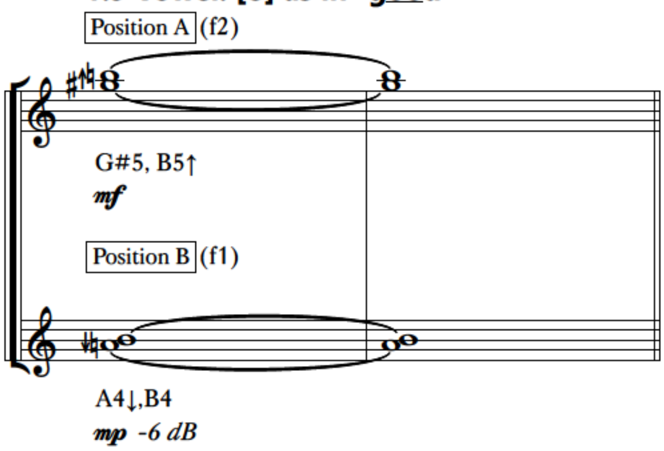
1.7 Vowel: [ɔ] as in "thaw"



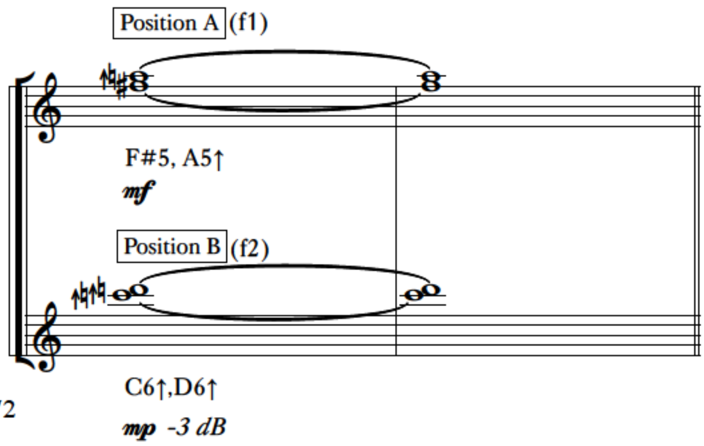
1.8. Vowel: [u] as in "mood"



1.9 Vowel: [ʊ] as in "good"



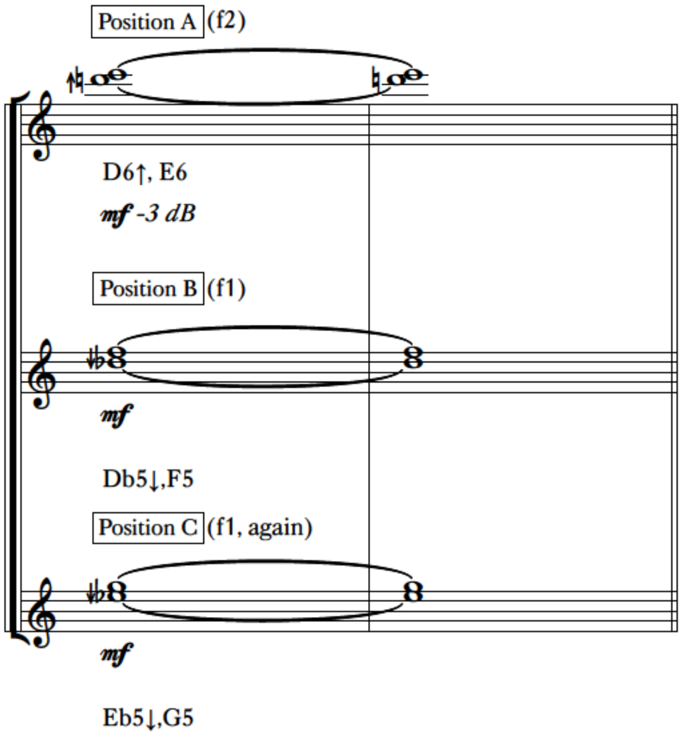
1.10. Vowel: [ʌ] as in "strut"



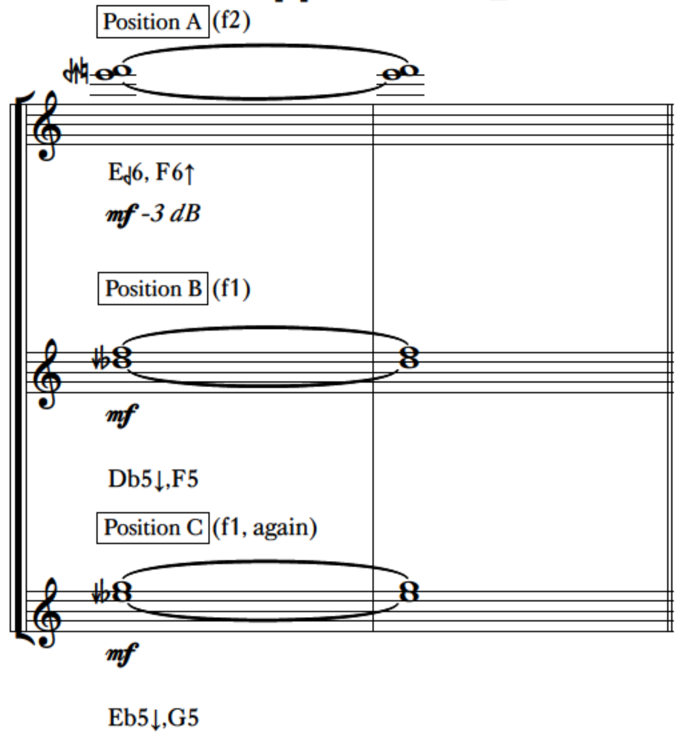
♩= 60

1. MONOPHTHONGS (3/3)

1.11. Vowel: [ɜ] as in "stern"



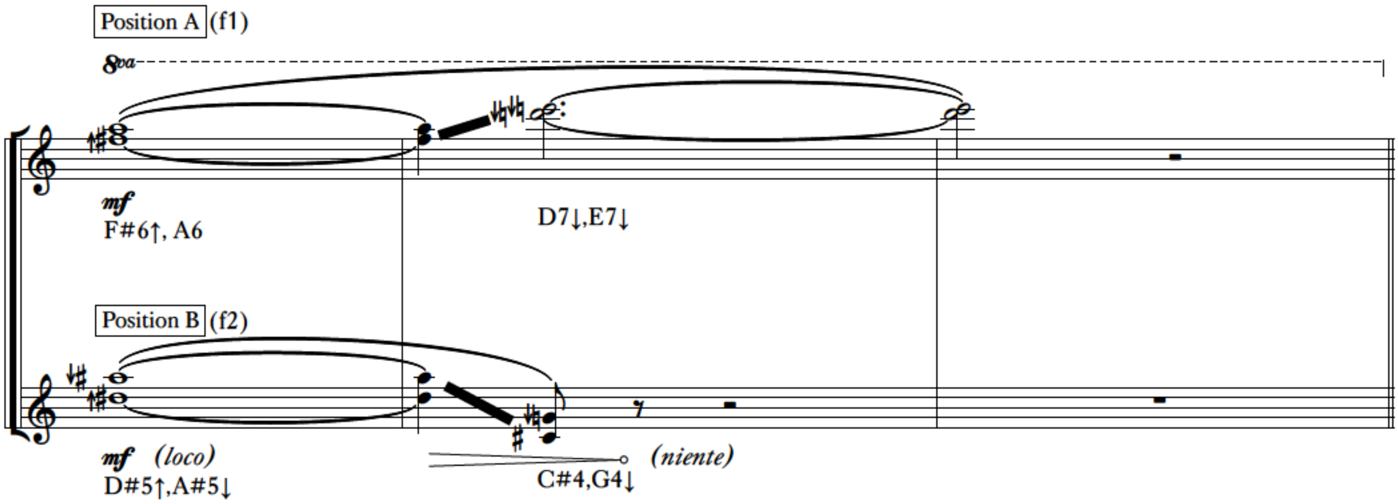
1.12. Vowel: [ə] as in "comma"



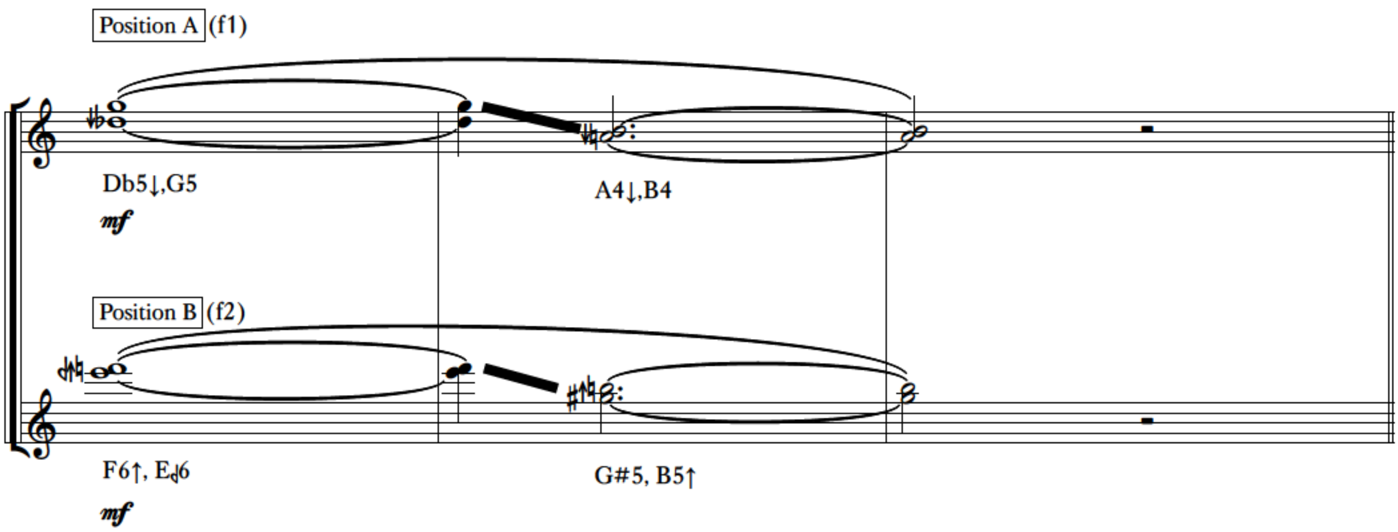
♩= 144

2. DIPHTHONGS (1/2)

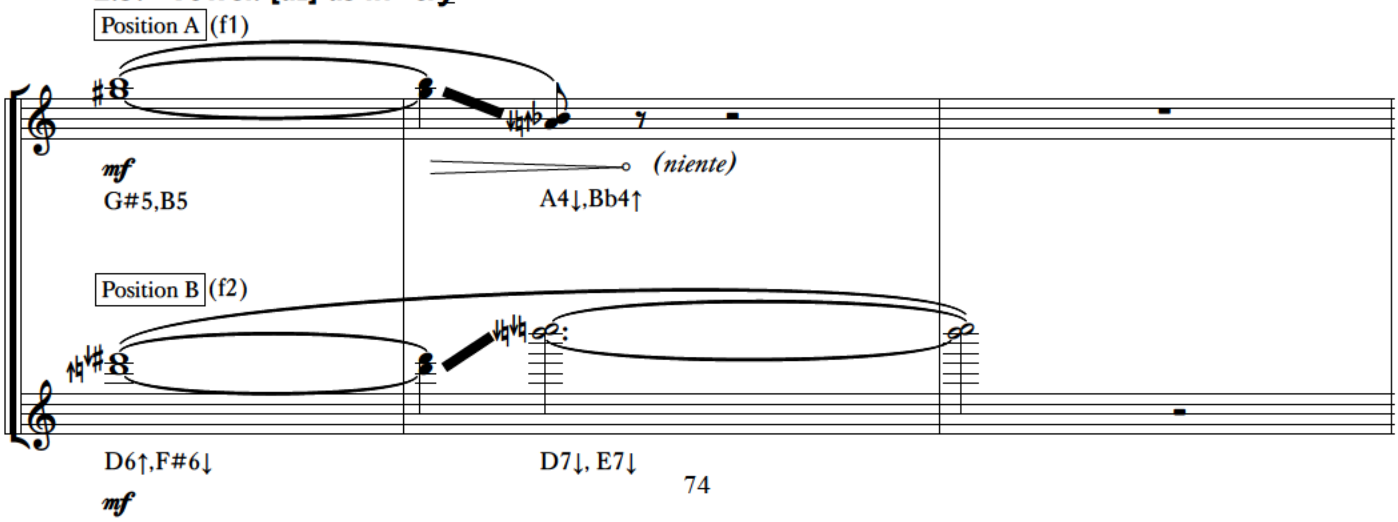
2.1. Vowel: [eɪ] as in "stay"



2.2. Vowel: [əʊ] as in "toad"



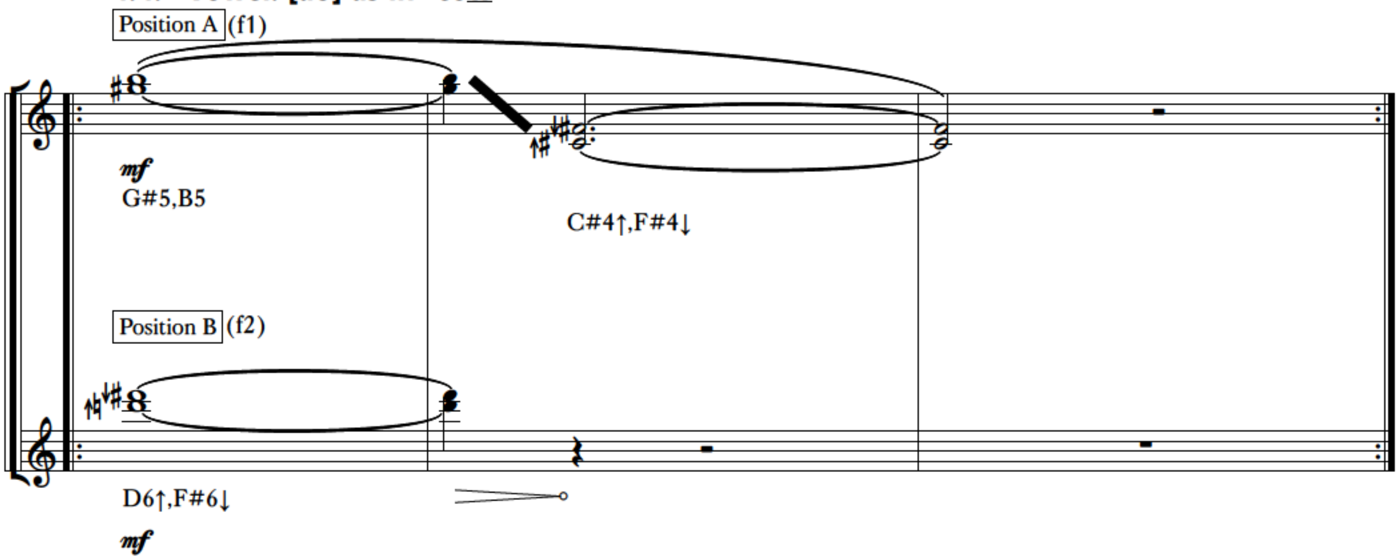
2.3. Vowel: [aɪ] as in "cry"



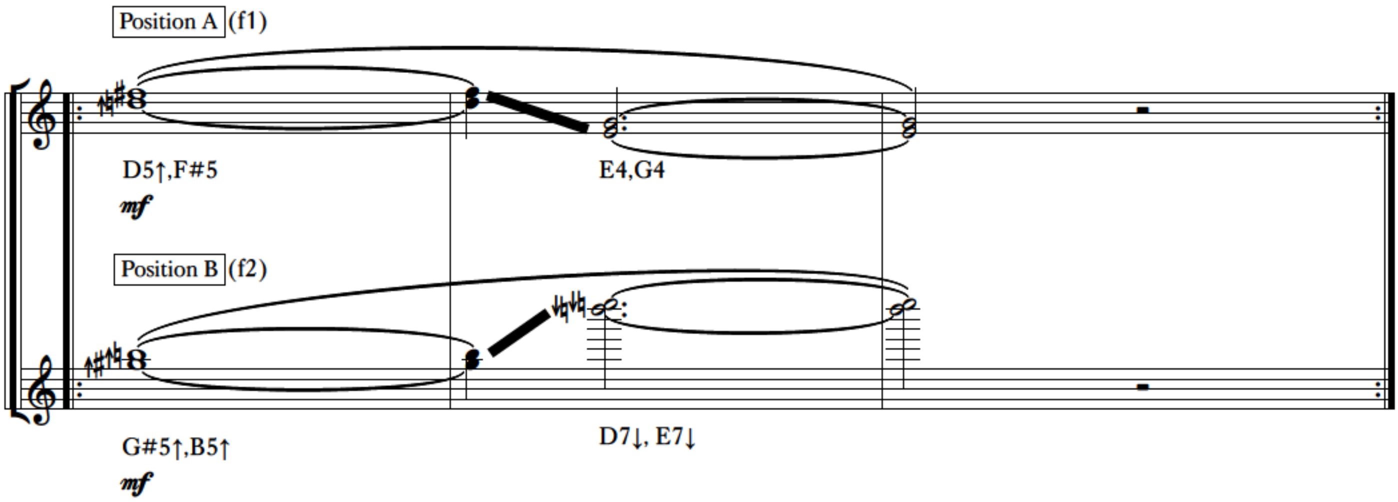
♩= 144

2. DIPTHONGS (2/2)

4.4. Vowel: [aʊ] as in "cow"



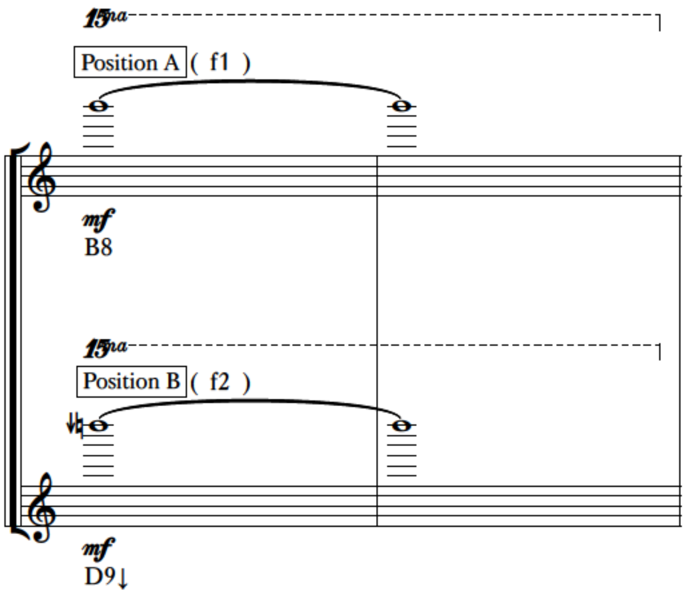
4.5. Vowel: [ɔɪ] as in "ploy"



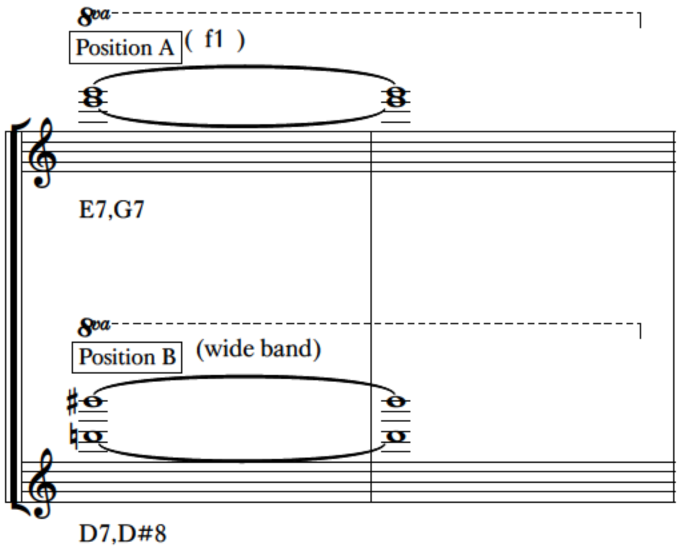
♩= 60

3. FRICATIVES AND SIBILANTS (1/1)

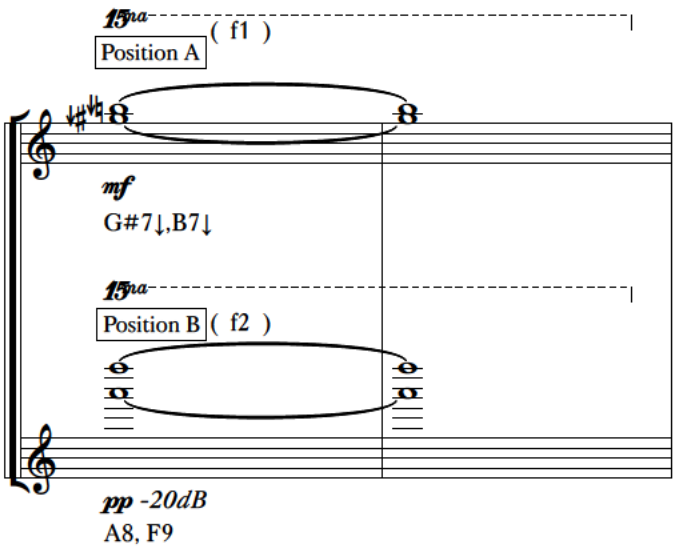
3.1. Vowel: [s] as in "sinew"



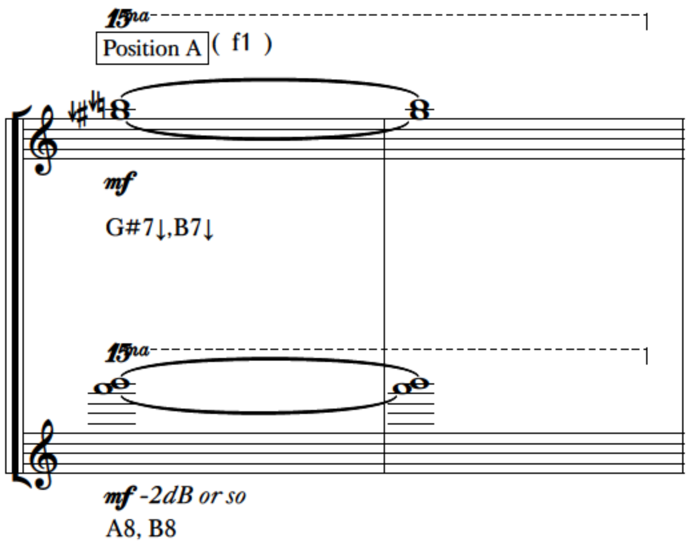
3.2. Vowel: [ʃ] as in "shake"



3.3. Vowel: [f] as in "fun"



3.4. Vowel: [θ] as in "thespian"



APPENDIX B: List of recorded phonemes and relevant timestamps

Label	Phoneme	Timestamp
1.1	[i]	0m 00s
1.2	[ɪ]	0m 11s
1.3	[ɛ]	0m 24s
1.4	[æ]	0m 39s
1.5	[ɑ]	0m 59s
1.6	[ɒ]	1m 07s
1.7	[ɔ]	1m 21s
1.8	[u]	1m 37s
1.9	[ʊ]	1m 52s
1.10	[ʌ]	2m 10s
1.11	[ɜ]	2m 25s
1.12	[ə]	2m 39s
2.1	[eɪ]	3m 12s
2.2	[əʊ]	3m 23s
2.3	[aɪ]	3m 32s
2.4	[aʊ]	3m 34s
2.5	[ɔɪ]	3m 55s
3.1	[s]	4m 20s
3.2	[ʃ]	4m 34s
3.3	[f]	4m 49s
3.4	[θ]	5m 05s