The Qualities and Flow of Imagined Sound and Music

Chris Chafe

Abstract A new lens on experience is proposed. It is a way of understanding how we produce rhythmic syncronization between each other. Our musical ability for fine ensemble precision is exquisite in the way it adapts, predicts and communicates. Studying this empirically takes at least two participants but understanding its mechanisms requires entering into the minds of individuals. There, in the realm of imagination, is where sound lives on a different kind of stage. Its churning temporal flows are laced with memory and anticipation, consuming the present and producing it, as well.

`Key Words:` Imaginary Sound, Music, Phenomenology, Psychology of Time.

1 Sound in the Imagination

When we imagine sounds and music our "tonal thoughts" have qualities like loudness or timbre which resemble the qualities which we ascribe to perceived sound. Or are these dimensions of imagined sound analogies, somehow borrowed, weaker, or imposed? Imagined sounds are not accessible for measurement like their counterpart physical phenomena. Qualities of perceived sound are graspable in part because we can link them to measurable attributes of external sounds. That which a microphone (or earlier pre-electronic apparatus) registers can be compared to that which we report via listening. The physical signal can be repeated, analyzed and described in physical dimensions. Links between physical properties of sound and related perceptual attributes are the product of centuries-long work in acoustics and psychoacoustics (Hui 2012).

I'm certain that the greater fraction of my auditory experience is, in fact, internal. Modes of hearing and listening vary continually – perceived, imagined, unconscious,

Chris Chafe
Center for Computer Research and Acoustics, Stanford University, e-mail: cc@ccrma.stanford.edu

conscious, reflex-triggering, enacting – and it's those which take place on the proscenium of the imagination which are the least studied. Alexandra Hui found it "curious this skill was not of psychophysical interest" in the time of Helmholtz, a time of the "geistige" ear and the theory of unconscious inference and when "The listener's ability to generate or at least reproduce sounds absent the stimulation of actual sound waves." was certainly apparent at the time (Hui 2012).

Most people are able to willfully imagine sounds. For example, to conjure the sound of a familiar musical instrument or to attend to their own inner voice while reading silently. Conscious manipulation of auditory imagery is also commonplace. Given phrases of text, participants can imagine them spoken by friends or relatives and will usually report that it's easy to do. The first of two pilot studies described below gave me a sense of just how ubiquitous this capability is. Of 100 subjects, 98 reported they could imagine the sound of "a hard ball dropped from waist height onto pavement." The Bucknell Auditory Imagery Survey (Halpern 2015) tests the vividness with which a participant can imagine sounds and change them. The survey presents a series of imagery tasks by textual description. Subjects are asked to imagine, for example "The sound of an all-children's choir singing the first verse of a song" and then "An all-adults' choir now sings the second verse of the song."

Why does this matter to me as someone mostly working in computer music? The answer begins with an entirely different line of inquiry. Using the Internet for high-quality "teleconcerts" and rehearsing Fig. 1 has been a research project which I've been involved with for two decades (Chafe 2000). The work has become increasingly important during the present COVID-19 pandemic. There is great interest on the part of musicians worldwide who are directly impacted by the impossibility of gathering in person for group music making. As soon as network music performance became technically feasible and first experiences showed promise for rhythmic syncronization, we tested pairs of subjects clapping together to learn about the effect of very short time delays on their ensemble tempo coordination (Chafe 2004)(Chafe 2010). Questions of interest in these studies included, "What is the latency limit (in msec) beyond which maintaining a shared pulse becomes difficult?" And, "If tempo degrades with increasing lag, are there naturally-occuring coping mechanisms which arise in ensemble playing?" The range of delays tested covered a large portion of the range experienced in network music performance situations (from 1ms to 78ms one-way).

Our early study, which was run between adjacent studios in our center, featured the simple clapping pattern shown in Fig. 2. The experiment was conducted using headphones with very little ambient reverberation and only a sparse set of starting tempi. Two of the recorded trials can be heard in these sound examples, Fig. 3 and Fig. 4. You can hear the sounds by using a mobile phone with a barcode reader and aiming the camera at the each figure. Open the link it provides in a browser. Any modern browser should be able to play the sounds. [1] The two recordings illustrate the effect of manipulating temporal separation when a pair of clappers are separated with minimal latency (15 ms) and very long latency (78 ms).

[1] I am indebted to Diana Deutsch for introducing this practical way to include sound examples in a book.

Those results led to modeling work in which two interacting "clapper algorithms" interact to approximate the same behavior. Perhaps it's no surprise that the early clapping models (Gurevich 2004) (Caceres 2013) under-performed relative to our clapping humans in terms of tempo stability as shown in Fig. 5. Our ability to synchronize and to adapt our music making to different conditions is superb. We live in and participate in event flows and the intriguing challenge now is to tease apart the how's and why's of this very human skill. Applying the latest in synchronization theory and innovating in the mathematical description of coupled synchronous processes has increased the competence of modeling algorithms (Roman 2019) but performers cope with these short time lags in ways we have yet to understand. This is where the interest in auditory imagery ties in. Can we experiment directly with rhythmic flows at the level of protention and retention? These are names for the mental near-time in which events are planned to happen and have just happened.

Fig. 1 A split ensemble with members playing together located in California, Michigan, New York, Belgium and Germany.

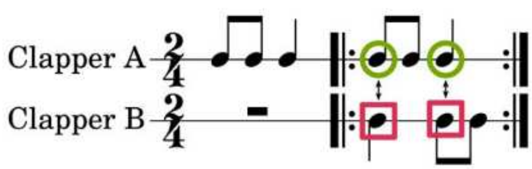

Fig. 2 The test rhythm. Pairs of people clapping the rhythm together were situated in separate rooms (Chafe 2010).

2 The Specious Present

As listeners and performers, we are good examples of creatures who live in the "specious present." The ways in which we think ahead in time, for example reading a bar ahead while performing from a score or anticipating upcoming chord changes in a tune or vaulting through the arch of a phrase we're building, these are all examples of planning and thinking in tones. William James (following Kelly, and followed by Husserl and others) contributed descriptions of temporal flow in the mind.

> The practically cognized present is no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched, and from which we look in two directions into time. The unit of composition of our perception of time is a duration, with a bow and a stern, as it were—a rearward- and a forward-looking end. (James 1890)

It seems likely that some of the difference between a mechanistic model using two interacting clapper algorithms and two human clappers will reside in how such flows are engaged in performance. In summarizing his modeling attempt, Juan-Pablo Caceres wrote:

> The model presented performs better than previous attempts, but humans are still better at this task. Second-order adaptation in a finer inter-beat granularity that are not accounted in the presented model may explain this. Our model includes an anticipation parameter that is static, and a better understanding of this prediction mechanism is a desirable goal to improve the performance of the model. (Caceres 2013)

The next step in creating a lens for examining attributes of human clappers' performance is to combine our increasingly sophisticated models with humans. Improvements in modeling are bringing us closer to that goal. In his work incorporating "strong anticipation" Iran Roman found that there may be a simpler approach.

Fig. 3 Recordings of three trials of the clapping rhythm shown in Fig. 2 when participants were 15 ms apart. You can hear the sounds by using a mobile phone with a barcode reader and aiming the camera at the figure. Open the link it provides in a browser. Any modern browser should be able to play the sound.

Fig. 4 Recording of three trials of clapping rhythm shown in Fig. 2 when participants were 78 ms apart.

> We have demonstrated that one can use a dynamical system with delayed feedback to simulate human anticipation during perception-action coordination. We only needed to add delayed feedback to an oscillator in order to explain human anticipation. A more complex mechanism of anticipation, like statistical inference, was not necessary. (Roman 2019) [2]

If we're nearly there and the algorithm / human subject combination can provide a means for rapid experimentation then we can broaden the experimental conditions. The ability to test, for example, a wider set of musical behaviors and different acoustical conditions, and then evaluate the results more rapidly is key to learning more about possible factors involved.

The present chapter presents preliminary aspects and aims at gaps which I see in our understanding of "thinking in sound and music." What is particularly relevant to performance is the our lack of understanding of "planning sounds" internally. The present discussion now moves from this narrower starting point of modeling to ask more widely about the content of auditory imagery. Temporally-related objects, but what objects? The arguments to be made will set the stage to move full circle back to modeling if, as seems likely, simple structures emerge as they have in related fields. Temporal "chunks" of sound have explanatory value in linguistic discourse (Chafe 1994)(Chafe 2018). Auditory stream formation of perceived sound has well-known importance in parsing complex sound worlds (Bregman 1990). "Socially-endowed internal models" have been proposed in which the performer internally simulates their co-performers' immediate tendencies (Wolpert 2003)(Keller 2008)(Keller 2012). Marc Leman suggests structures related to "imagery and object concepts" in his discussion of sensorimotor prediction (Leman 2016).

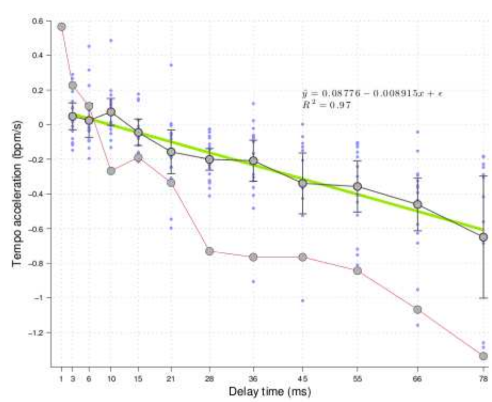

Fig. 5 All trials of all pairs of people clapping the rhythm together shown above in Fig. 2. Averages of tempo deacceleration vs. temporal separation (black) with linear regression (green). The same task performed by a computer model with human syncronization traits (red) (Caceres 2013).

[2] "Strong anticipation" is explained by Roman. "Specifically, 'anticipatory synchronization' in strong anticipation emerges from the coupling between a 'response' system and a 'driver' system (e.g., stimulus input) wherein the response system also receives delayed feedback about its own activity. One of the major strengths of the strong anticipation approach is that it accounts for anticipatory phenomena beyond human behavior, and that collectively all such phenomena can be modeled as coupled dynamical systems."

Brain studies using imaging techniques (Halpern 2004)(Herholz 2012) and EEG observations (Schaefer 2011)(Schaefer 2013) are advancing correlations with auditory imagery. If we call that a kind of "outside-in" experimentation, what I propose here is on a parallel front, from the inside-out. It involves gathering introspective evidence (Casey 2000)(Ihde 2007) to describe aspects of temporal flow (Schmicking 2005)(Varela 1999) and tonal flows (Hodges 2011)(Keller 2006). The following pilot studies demonstrate how self-reports can be provide such evidence with the advantage that today such surveys can reach an internet-enabled, scaled-up population of "imagers." Eventually, these same techniques can be used with musician subjects. Talks which I've given on the subject have titles which develope the flavor of the approach: "The Acoustics of Imagined Sound" and "The Sound Stage of the Mind: Imagined Sounds and Inner Voices, Probing inner sound via crowd-sourced self reports." The seeds are in the ground as far as motivation and proof-of-concept and what remains now is to adapt the techniques to questions related to timescape and flow.

2.1 Pilot Studies (2013-14)

The pilot studies asked questions aimed at describing qualities of mainly static mental objects. Study 1 surveyed 100 online "workers" for clarity ratings of their own inner voice, friends' voices, musical instruments and environmental sounds. The rating scale was adopted from the Vividness of Visual Imagery Questionnaire (VVIQ Scale) (Marks 1973). As an example, Fig. 6 shows the results for 100 subjects rating the clarity of the imagined sound of "a hard ball dropped from waist height onto pavement."

In addition to gathering ratings of clarity of certain imagined sounds, the survey probed imaginary acoustical dimensions of location and relative loudness. Fig. 7 shows the results when subjects were asked to silently read the words "mechanical turk" and judge whether the sound of their inner voice was located inside or outside their head. Fig. 8 shows which imagined sound is louder when comparing a few suggested sounds. These examples were only intended to test the efficacy of

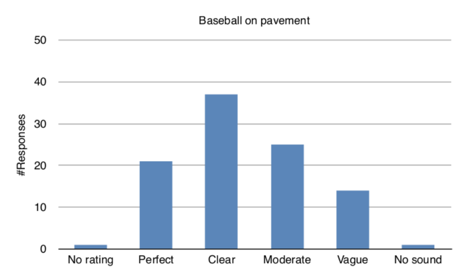

Fig. 6 Pilot study 1: Vividness self-reports of 100 subjects rating the clarity of the imagined sound when imagining the sound of a hard ball dropped from waist height onto pavement.

crowd-sourcing judgements about imaginary sound made through self-reports. More methodical manipulations followed in the next experiment.

Study 2 (50 subjects) involved listening to sounds as well as imagining them. It began by introducing a set of novel synthesized sounds played from the browser and then experimented with imaginary recall of those sounds later in the survey. You can hear the sounds using the same technique as above. As will be heard, the set of sounds increases in loudness Fig. 9. They were generated with a novel physical model which synthesized something resembling a bowed metal plate and their intensities were calibrated to increase in 6 dB steps. Subjects were presented with them in random order and were asked to rank whether they were louder or softer than a recorded voice. The unfamiliar sound and the voice probe could be repeated as needed by pushing buttons on the browser interface.

Later in the survey, subjects were asked to recall the voice and recall the unfamiliar sounds across their range from softest to loudest. The recalled sounds were again compared for loudness and these answers provided rankings which allowed for comparison of the two sets, namely perceived and imagined. Two loudness curves were produced which are shown superimposed in Fig. 10. The imaginary one, which is the one of interest, appears to be a compressed version of the perceived one. This finding is consistent with the theory of imaginary extension (Casey 2000). Briefly stated, it's the phenomenon that imagined horizons are limited in their extent and that the distances between in which imaginary scenes are constructed have a relatively smaller scale than the same thing perceived. In terms of loudness, imaginary soft is not as perceived soft and loud is not as loud.

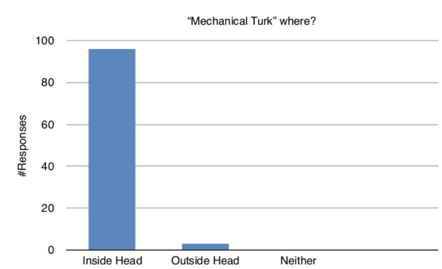

Fig. 7 Pilot study 1: Self-reports of 100 subjects judging whether the sound of their inner voice was located inside or outside their head when subjects asked to silently read the words "mechanical turk."

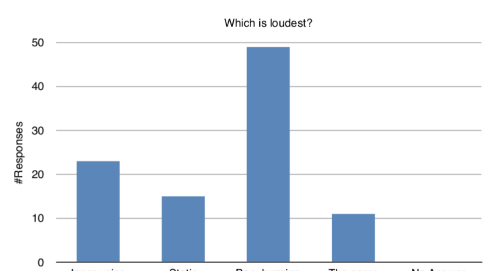

Fig. 8 Pilot study 1: Self-reports of 100 subjects judging which of three imagined sound was loudest (their inner voice, static, or a bee buzzing).

3 About Imaginary Loudness

Daniel Schmicking poses the question “Is there imaginary loudness?” in an essay on phenomenological method (Schmicking 2005). The article has merit for its critique of approaches and because its conclusions suggest openings for further investigation. He proposes that “agreement procedures” could be developed to confront the former. Introspective evidence is difficult to validate, otherwise. “Cooperating phenomenologists then should be able to decide whether there is always quasi-loudness in auditory imagery, even if they could not claim apodicticity.” He makes a strong assertion that there is indeed imaginary loudness (quasi-loudness).

> If you doubt the quasi-loudness of your own inner speech or voice just ask yourself: does the voice whisper or does it roar? Does it sound like calling from a great distance, from the room next door, or rather from within your head?
>
> You are very likely to (implicitly) imagine accents along with the melody. This is the reason why the accentuated notes are quasi-louder; otherwise your melody, lacking accents, will be rhythmically ambiguous.

These examples ultimately require greater precision about loudness itself. From our perceived impressions of the physical world we know that confounding cues like timbre and attack, and underlying concomitants like force need to be accounted for. Impressions are not unidimensional and illusions abound. Can the whisper itself

Fig. 9 Pilot Study2: The five synthesized sounds which subjects played from a browser for experimentation with later imaginary recall during the survey. You can hear the sounds using the same technique as above. As will be heard, the set of sounds increases in loudness.

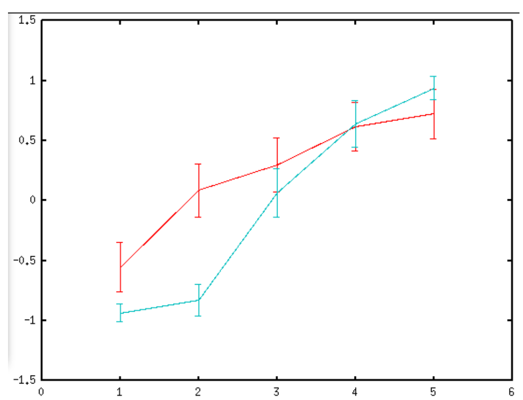

Fig. 10 Pilot Study2: Compared to the perceived scale (blue), the imaginary quasi-loudness scale (red) is compressed and shifted.

can be made louder? If it can, what's the "quasi" mechanism producing the change? Does it get louder by whispering more (quasi-)forcefully or by (quasi-)amplification of sound pressure level? Many other ways exist to make a whisper louder, such as cupping your hands or moving closer. Informally, in discussions of self-reports of imaginary loudness, I've noted a prevalence for force being strongly coupled with impressions of loudness and less so that sound pressure level is coupled with imaginary loudness. That's because it seems to be more difficult to perform strictly intensity-based manipulation in your mind. It's apparently easier to imagine hitting an object harder than to imagine the sound of a hit emitted via a device with a variable volume control. But some subjects who report they can accomplish this often report that the task evokes an image of a volume knob on an amplifier or volume slider in a sound playback application. Imaginary context helps.

As Schmicking himself points out, it's impossible to know whether one is instead imagining a difference based on distance.

> But in like manner an audio-imaginary (quasi-auditory) scene is related to the implicit imaginary body. When this perspectival organization of perception is transmitted to imagery then there might as well be a quasi-distance of imagined sounds and hence a correlative quasi-loudness. The quasi-loudness of imaginary events varies with imagined changes of quasi-distance.

Imaginary accent as a proof of the existence of imaginary loudness also requires closer examination since accent can depend on relative pitch height, timbre, attack and timing. My first pilot study (which compared only static mental objects, not melodies) found (quasi-)timbral manipulation was easily accomplished across a large number of subjects.

In conclusion, Schmicking leaves the question open.

> Perhaps this might still be contested since it may be arguable that not every subject is able to focus on imagined loudness; or maybe there 'is' loudness of imaginary sounds only when one's attention is shifted to it by external factors (e.g., an experimenter's instructions), or there may even be subjects who cannot imagine loudness at all.

4 Constructing the New Lens

Schmicking's challenge here is apt and the pilot studies were an initial attempt to use crowd-sourced surveys to gain statistical significance. The results were encouraging enough that I can propose extending the technique toward questions of temporal flow. I intend to further engage crowd-sourcing services (like Amazon's Mechanical Turk which was used in the pilot studies) in order to reach large numbers of dedicated survey takers. The goal is to test "phenomenological agreement" of a statistical kind with little constraint on the size of subject pool. In these studies as in the pilot experiments, browser-based presentation of sounds and music, and inner voice will figure in the designs. One possibility is to engage mental counting as a way of accessing and "time-stamping" events and objects in the "near now."

These are some experimental design criteria for going forward:

1. accounting for subject variability,
2. eliminating experimenter bias,
3. methodically entering into the realm of the “specious present” and
4. enhancing the validity of online surveys.

There are prior studies that inform strategies for each.

1. The BAIS will be useful for characterizing participants’ individual aptitudes.
2. Gelding, et al. have devised an experiment for isolating imaginary pitch with a novel degree of purity (Gelding 2015).
3. Pecenka (Percenka 2013), Keller (Keller 2006) and others have shown a relationship between musician’s auditory imagery abilities, temporal prediction and sensorimotor synchronization. In one case, participants were asked to to mentally continue a tempo change across a short auditory sequence with a gap, and then to judge whether a probe tone occurred early or late relative to the imagined continuation.
4. Validity of online surveys has been a concern in the social sciences and methods for reinforcing their significance are available to be put into practice (Berinsky 2014).

The designs will attempt to characterize mental sound objects, object “chunking” and temporal flow with a particular goal of examining the near-time “retention / protention” qualities described by Husserl. Husserl’s attempts to diagram this mental timescape are fascinating, Fig. 11 and Fig. 12(Dodd 2005). They depict event objects in procession as they recede from now into the near-past and as they set up expectations in the near-future. If one were to diagram echoic memory in which the persisitence of the sound of events gives way from veridical to schematic in about a half-dozen seconds, the result might resemble these sketches of Husserl’s.

The physical timescape of musical syncronization at the micro-time level was studied in our work with pairs of subjects clapping together(Chafe 2010). A prototype system has been tested in which an adaptive and listening rhythm tapping algorithm is coupled with a human tapping on a keyboard, Fig. 13. Leading and lagging in very small proportions is visible. Keeping a steady pulse is a dynamic process with alternating give and take. The braided appearance is typical of duos being analyzed at this level of temporal resolution. Combining the ideas present in Husserl’s timescapes with micro-time measurements will give us a new lens on imaginary flows of sound and music.

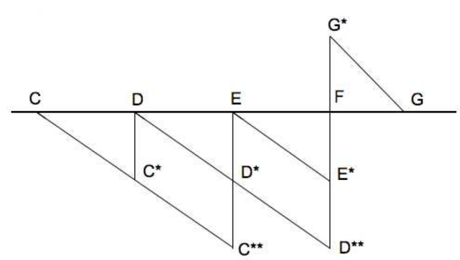

Fig. 11 Husserl’s time chart shows it as an assemblage of experienced objects.

5 Aspects of the Proposed Research

Imaginary sound as a field of research is expanding. Scientific and artistic projects are motivated by a curiosity to understand more about this hidden realm and manipulate its qualities. It's clearly a central element of human experience. For my part, my interest begins with my seemingly tangential research and practice involving online music performance. It was there that I discovered behaviors and paradoxes that led me to consider related studies of time consciousness.

Three components make up the further study proposed by the present contribution: motivation, literature and method.

The first is to give an account of the fundamental inverse relationship between increasing temporal separation and tempo. Surprisingly, our data showed that there was a range of extremely short delays with tempo acceleration. The finding supported the notion of a natural delay range for best tempo stability on the order of delays in familiar acousical settings. The "sweet spot" is similar to the propogation delays in air between musicians in rooms and on stages. For tight rhythmic syncronization in network music performance it's especially important to replicate this range in engineering the online systems. What of these accelerating flows? These are manifest in a context requiring two clappers. Are human-human coupled flows one flow or two? Will algorithm-human coupled flows be a way into manipulating that? Methodically observing qualities of such flows with self-reports has not yet been tried.

Further study will include more deeply tracing the topic historically across domains. Literature on temporality spans philosophy, linguistics, psychophysics and music. The data include everything from solo arm-chair introspection through how temporal concepts are expressed in the world's languages to neuroimaging-based approaches. On the music side, there are a growing number of composers who manipulate the "sound thoughts" of their audience. Their compositions ask their listener participants to actively engage in auditory imagination. For example, the works of Pauline Oliveros, Vanessa Tomlinson and Amnon Wolman literally are played on the internal proscenium and are not static scenes. How do they engage temporal flow? There may be methods in these works which can inform experimental designs.

Lastly, I suggest that further studies take a cue from Schmicking (above) and examine quantitative methods in phenomenology. A useful product will be the establishment of a protocol which can be shared across experiments. Reproducibility

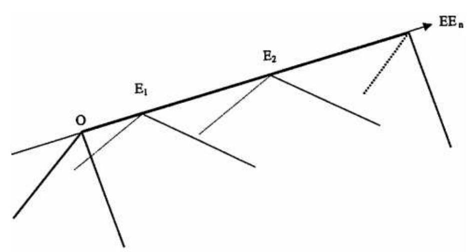

Fig. 12 Husserl's representation of objects with independent dimensions.

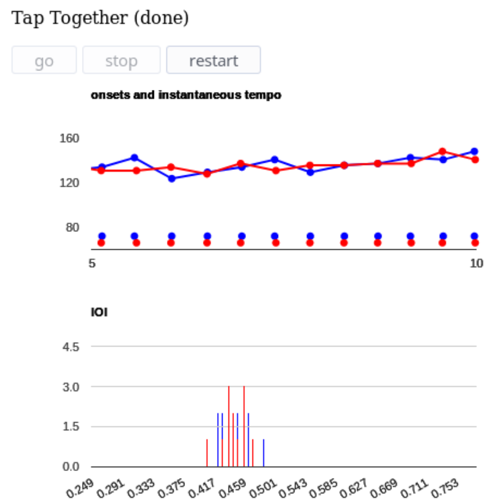

Fig. 13 Automatic tapping algorithm (red) and human tapper (blue). The inter-onset interval time series is plotted as a function of temp (above), time of occurence (middle) and histogram (below).

is essential, and providing a practical "howto" for crowd-sourced methods will allow replication of the findings.

Acknowledgements With thanks to Juan Pablo Cáceres, Hongchan Choi, Rob Hamilton, John Granzow, June Holtz, Pauline Oliveros, Andrea Halpern, David Huron, Chryssie Nanou, Jonathan Berger, Jieun Oh, Andy Stuhl, my Music 220 classes and many others who have contributed to discussions of imaginary sound, network music performance and techniques for crowd-sourced studies.

Appendix

As a part of both of the two pilot studies, survey-takers were asked for their own observations about the subject of imaginary sound. Their answers were revealing and a sampling of them is included here. Some encountered their inner voice for the first time, others experienced difficulties. The term "hit" refers to a task which Amazon Turk workers choose to complete.

1. This was a very interesting hit. It is one that really makes you listen and concentrate, thanks.
2. Is there any way to find out what this study is for? This was interesting to do.
3. This was an interesting task. I've never been asked to do anything quite like it before. This was fun to complete.
4. I found it interesting to not be able to imagine my voice coming from a spot outside myself or any voice saying mechanical turk including that of my mother. I could hear her voice in my mind but could not imagine or hear her saying mechanical turk. Odd and I wonder why that is.

5. very interesting study... a subject I've never thought about but makes complete sense.
6. Definitely a different type survey. It was fun. Thanks for being creative!
7. This was a weird but interesting hit. I had never thought about this aspect of hearing or inner voices. I will work with it a bit and see what happens. I realized while doing it than depending on where my focus is, I can hear the sound/voice in different manners- in my head- my voice or a different voice, outside my head like someone else is speaking, in my head like someone else is speaking, in my head with the voice I usually hear speaking to me- whatever I think I can make physically happen.
8. As I began the task, I was able to imagine hearing my voice. Then as the task went on, it became more difficult to hear anything, because there was no sound.
9. This was interesting. I have very good hearing and often hear things other people don't. However, I wasn't very good at imagining a friend's voice or a sound coming from outside my head.
10. This was an unusual but very interesting study. It is almost like separating body and soul.

References

Bregman, Albert. 1990. *Auditory scene analysis.* Cambridge, MA: MIT Press.

Berinsky, Adam J., Michele F. Margolis and Michael W. Sances. 2014. Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American J. of Political Science.* doi: 10.1111/ajps.12081.

Bregman, Albert. 1990. *Auditory scene analysis.* Cambridge, MA: MIT Press.

Cáceres, Juan-Pablo. 2013. *Synchronization in rhythmic performance with delay.* Stanford, CA: thesis.

Casey, Edward. 2000. *Imagining: a phenomenological study.* Bloomington, IN: Indiana University Press.

Chafe, Chris, Scott Wilson, Randal Leistikow, Dave Chisholm and Gary Scavone. 2000. A simplified approach to high quality music and sound over IP. COST-G6 Conference on Digital Audio Effects. Available via DAFX Conference Web Page. https://www.dafx.de/paper-archive/2000/pdf/chafe.pdf. Cited 19 Sep 2020

Chafe, Chris, Michael Gurevich, Grace Leslie and Sean Tyan. 2004. Effect of time delay on ensemble accuracy. Proc. of the International Symposium on Musical Acoustics 31 Available via CCRMA Stanford University.
https://ccrma.stanford.edu/ cc/pub/pdf/ensAcc.pdf. Cited 19 Sep 2020

Chafe, Chris, Juan-Pablo Cáceres, and Michael Gurevich. 2010. Effect of temporal separation on synchronization in rhythmic performance. *Perception.* doi: 10.1068/p6465.

Chafe, Wallace. 1994. *Discourse, consciousness and time.* Chicago: Univ. of Chicago Press.

Chafe, Wallace. 2018. *Thought-based linguistics: How languages turn thoughts into sounds.* Cambridge: Cambridge University Press.

Dodd, James. 2005. Reading Husserl's time-diagrams from 1917/18. *Husserl Studies.* doi: 10.1007/s10743-005-6403-2.

Gelding, Rebecca W., William F. Thompson and Blake W. Johnson. 2015. The pitch imagery arrow task: effects of musical training, vividness, and mental control. *PLoS ONE.* doi: 10.1371/journal.pone.0121809.

Gurevich, Michael, Chris Chafe, Grace Leslie and Sean Tyan. 2004. Simulation of networked ensemble performance with varying time delays: Characterization of ensemble accuracy. Available via CCRMA Stanford University. https://ccrma.stanford.edu/ cc/pub/pdf/simNetEnsPerf.pdf. Cited 21 Sep 2020

Halpern, Andrea. 2015. Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain.* doi: 10.1037/pmu0000081.

Halpern, Andrea R., Robert J. Zatorre, Marc Bouffard, Jennifer A. Johnson. 2004. Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia.* doi: 10.1016/j.neuropsychologia.2003.12.017.

Herholz, Sibylle C., Andrea R. Halpern, Robert J. Zatorre. 2012. Neuronal correlates of perception, imagery,and memory for familiar tunes. *J. of Cognitive Neuroscience.* doi: 10.1162/jocn_a_00216.

Hodges, Donald A. and David C. Sebald. 2011. *Music in the human experience: An introduction to music psychology.* New York: Routledge.

Hui, Alexandra. 2012. *The Psychophysical ear: Musical experiments, experimental sounds, 1840-1910.* Cambridge, MA: MIT Press.

Husserl, Edmund. 1928. *Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins.* Tübingen: Max Niemeyer Verlag.

Ihde, Don. 2007. *Listening and voice: Phenomenologies of sound.* Albany: State University of New York Press.

James, William. 1890. *The Principles of pschology.* New York: Holt.

Keller, Peter. 2008. Joint action in music performance. In *Enacting intersubjectivity: A Cognitive and social perspective to the study of interactions*, eds. Francesca Morganti, Antonella Carassa and Giuseppe Riva 205–221. Amsterdam: IOS Press.

Keller, Peter E. 2012. Mental imagery in music performance: underlying mechanisms and potential benefits. *Annals of the New York Academy of Sciences.* doi: 10.1111/j.1749-6632.2011.06439.x

Keller, Peter and Iring Koch. 2006. The planning and execution of short auditory sequences. *Psychonomic Bulletin & Review.* doi: 10.3758/BF03193985.

Leman, Marc. 2016. *The Expressive moment: How interaction (with music) shapes human empowerment.* Cambridge, MA: MIT Press.

Marks, D.F. 1973. Visual imagery differences and eye movements in the recall of pictures. *Perception & Psychophysics.* doi: 10.3758/BF03211175.

Percenka, Nadine, Annerose Engel and Peter E. Keller. 2013. Neural correlates of auditory temporal predictions during sensorimotor synchronization. *frontiers in Human Neuroscience.* doi: 10.3389/fnhum.2013.00380

Roman, Iran R., Auriel Washburn, Edward W. Large, Chris Chafe, Takako Fujioka. 2019. Delayed feedback embedded in perception-action coordination cycles results in anticipation behavior during synchronized rhythmic action: A Dynamical systems approach. *PLOS Computational Biology*. doi: doi.org/10.1371/journal.pcbi.1007371

Ross, Stewart L. 1985. The effectiveness of mental practice in improving the performance of college trombonists. *J. of Research in Music Education*. doi: 10.2307/3345249

Schaefer, Rebecca S., Rutger J. Vlek and Peter Desain. 2011. Music perception and imagery in EEG: Alpha band effects of task and stimulus. *International J. of Psychophysiology*. doi: 10.1016/j.ijpsycho.2011.09.007

Schaefer, Rebecca S., Peter Desain and Jason Farquhar. 2013. Shared processing of perception and imagery of music in decomposed EEG. *NeuroImage*. doi: 10.1016/j.neuroimage.2012.12.064

Schmicking, Daniel. 2005. Is there imaginary loudness? Reconsidering phenomenological method. *Phenomenology and the Cognitive Sciences*. doi: 10.1007/s11097-005-7597-7

Varela, Francisco. 1999. The specious present: The neurophenomenology of time consciousness. In *Naturalizing phenomenology*, eds. Jean Petitot, Francisco. J. Varela, Bernard Pachoud and Jean-Michel Roy 266–314. Stanford, CA: Stanford University Press.

Wolpert, Daniel M., Kenji Doya and Mitsuo Kawato. A unifying computational framework for motor control and social interaction. 2003. *Philosophical Transactions of the Royal Society B*. doi: 10.1098/rstb.2002.1238