Space and Sound: a two-factor approach to pitch perception

Adam Morton – adam.morton@ubc.ca

 version 2.4 – March 2015

"He [Sibelius] once baffled a group of Finnish students by giving a lecture on the overtone series of a meadow." Ross 2007

abstract: I identify two components in the perception of musical pitches. Distinguishing them makes it easier to learn to recognize notes and keys. To back up this implausible claim I give a way of learning to identify the components separately. I also make some connections with the psychology of music and the philosophy of perception.

Contents

1. the project & the claims

I have been trying to find new ways of describing the ways musical pitches sound, and investigating whether having these descriptions feeds back to perception. A practical phenomenology (Dennett 1991.) The project may not be that unusual, among half-competent musicians and musically obsessed philosophers, and during several years I would often think idly about it, and do not very successful exercises with tuning forks. (I became good at unconsciously estimating which fork it was from a particular set, by its weight!) I devoted the summer of 2014 to the project and after producing a preliminary version of this essay continued to tinker with it in the fall of 014. Sometime in the early summer I stumbled on the spread/wobble distinction below, and from then on focused on ways of applying it more accurately and more quickly. Now I find that I can classify pitches, independently of their register, into several rough overlapping classes. I often have intuitions now about what key a passage of music is in: though I am sometimes very wrong I am quite often right. And though I make mistakes, sometimes grotesque ones, I can often classify a pitch within a tone or so. (Success varies with timbre, duration, and whether the sound is coming through stereo headphones or drifting through a window.)

The method I describe has raised me from a very low level of performance at pitch and key recognition to a mediocre one[1]. That's progress! (But this does raise the question of what it can do for someone beginning from a higher level.) My results may not be different from those that others use to similar ends, for example by knowing their approximate singing range. (Though I think it is potentially more accurate.) But it is interesting. And it leads to curious and interesting experiences, in which one hears tones as having the two factors I will describe. See appendix III for a dramatic such experience; see also section **10**. In terms of these, I can begin to say how different pitches sound different, and to use the descriptions to remember them. Continuing the list, though these are all aspects of the same few

1 I shall describe myself as a musical incompetent. In fact, I am a curious patchwork of capacities and failings. But the details are probably irrelevant.

realisations, I much more readily now hear sounds outside a musical context as having a determinate pitch, even when I cannot name it. When sounds are presented as music, I often now have definite impressions of the pitch of a note or the key of a passage, even when I cannot force this into the standard language of lettered notes. I find complex music less baffling, though I often cannot say why, and when listening to music in which patterns of modulation are essential I feel lost less often, even when I have little confidence that I am identifying the key changes correctly. (The analogy with colour below might help give a sense of this.) A moral might be that a tenuous hold on the advanced can help with the elementary.

Besides these practical effects there are interesting issues of perception and language, of describing sounds in novel ways. This essay describes what I have learned and what I have become able to do, in *roughly* the order I learned them. But I recommend reading through it all, including the appendices, before trying it. If I am right, I have discovered a basic aspect of pitch perception, and a technique that allows me to exploit it. Individual history and perceptions vary, though, and you may have to experiment to apply these ideas to your own case. So what I am presenting has three aspects. It is a description of a method, an account of my own history finding and using it, and a hypothetical theory that might connect the method with our hearing of pitches and locations. These should be separable.

I am choosing my words to prevent you from taking my claims either as trivial or as stronger than they are. An analogy with colour perception may help. (But it is only an analogy. Some readers of the first version said how interesting they found the colour/pitch link. But though synaesthetic phenomena are familiar, if very rare, they do not help anyone to whom they do not come naturally to learn or improve anything. Take the analogy just as a way of clarifying my intentions.)

Suppose that our thinking about colour was like our language and awareness of pitch. (I am not denying that there are fundamental reasons why it is not.) There would be a very small minority of super-acute chromatic sensitives who could confidently say "that fabric is just a little more blue than my mother's car, which she sold five years ago", or "if you use the colour on this sample you'll find when you get home and do a whole wall that it is just a little bit lighter than its neighbour." Most of these people would have known from an early age that they could do this. Meanwhile, the rest of us would stumble along with "bright", "vivid", and "now that I see them side by side they look very different." No doubt some people are much sharper about colour than others, but most of us are in neither of the two categories. We have vague useful words like "red", "light brown" and "mauve", of varying degrees of precision. And we can roughly compare present and remembered hues in a way that accords with the limits of our verbal precision. Children have to learn to see colours in accordance with the colour language of their elders, and when they are learning the vocabulary they make mistakes that seem bizarre to adults (Davidoff and Fagot 2010.).

We would also find, though this would be known only to experts and seem surprising to everyone, that most people represent precise hues more exactly than they can consciously describe them. If people were forced to 'guess' at colour matches of, say, their childhood clothes or the walls of rooms they visited briefly, they would do this more reliably and accurately than they would have predicted. But when asked to say in words what colours these clothes and walls were, they would perform much less well. (I predict that something like this is actually true with level of illumination and degree of saturation.) These facts would suggest that a more acute classification lurks in the perceptual apparatus, lacking a procedure to bring it to language and consciousness.

The analogy with colour perception can be taken further. We see colours by using retinal cone cells and associated neural systems sensitive to three overlapping ranges of wavelengths. (It would not work if there were only one: two can do it and do in some animals including a few humans. Ali & Klyne 1985.) The analysis and method I am describing postulates two sensitivities. They are combined to get pitch judgements, which are usually unconscious and not available to language. But each of them can be made and named separately. Then they can be combined in a process of conscious reasoning. By ordinary people! That is the central and obviously controversial claim. (Anyone who agrees with me too quickly is naïve: if the idea is right it is surprising that it is right.)

Another way of making the claim. Distinguish access from accuracy. People with absolute pitch vary in how precisely they can discriminate. With many there is a definite element of categorical perception: notes are assimilated to a fixed number of labels, usually the standard twelve semitones (Deutsch 2013b). So as a limiting case imagine someone who can verbalize, consciously access, whether a note is within a half-octave centred on B. They know when it is A, A#, B, C, C#, or D and when alternatively it is D#, E, F, F#, G, or G#, presumably with a region of uncertainty at the boundaries between the two (so D and G# would elicit less certain judgements). These two regions are essentially what I will later call the growing and shrinking categories. Such a person would be able to do things that most people cannot – they would have conscious access to roughly absolute pitch information – but they would not have a very accurate sense of pitches. My aim is to develop access, letting accuracy fall as it may. (A conjecture: accuracy can be refined, but access requires special gifts or an unusual approach. Given access and hard work accuracy will improve at least somewhat.)

We do have vague terms for timbres and pitch height. We say "sounds like a clarinet though it is too low; perhaps it is a basset-horn". And there are aspects of colour and illumination that only the talented and practiced can master. (In the old days of photography there were a few who didn't need a light metre; I suspect most people are incurably bad at separating saturation from brightness.) So my aim of developing more informative but still vague classifications of pitches might be pushing at a firmly locked door, neurally closed for most of us. I am sure that most psychologists of music, and most musicians who have tried to emulate their colleagues with absolute pitch, would think that this is the case. But I think I have a way of doing it. This essay describes what I think I have stumbled on, late enough in life that my hearing has dulled and my nervous system is no longer plastic. I think that the reason many well-informed people consider the task impossible is that they take the aim to be to emulate perfection rather than to make a better accommodation to what one can and cannot do. (And, also, I suspect, they are reacting to fraudulent or deluded schemes for developing in many what only a few are capable of.)

For an analogous case with olfaction, see Majid and Burenhult (2014).

2. the components

Begin with the location of sounds. In everyday life this is dominated by the difference in loudness between the two ears, and subtle features of the way one's outer ear reflects sound from different directions, but I am sure these are not the factors that matter for my purposes. Listen to a pure tone, close your eyes so you do not see the source and ask yourself "if it wasn't where it seems most likely to be, what would be my second guess about location?". (Stereo headphones are good for this.) If you were to grab at it in the dark where would you move your hands and where would you be prepared to grab next if there is nothing where you first move? You'll find that the sound has several locations, and the purer the note the more distinct they are. You'll find they move. If you don't hear the multiple locations right away try forcing the sound to move, and note what displacements are easy and which take effort. (Forcing it can be treacherous at a later stage; I suggest this to get you into the wandering pure tones mentality.) You *may* find it is more vivid if you ask for the direction of the sound given the additional information that it is behind rather than in front of you. (For the wealth of information we use to locate sounds see Bregman 1990, 658-660.)

Don't invest time in honing this awareness. There are several things going on and it is important to separate them. (I lost months not realizing this fact.) Distinguish two factors, which I shall call *spread* and w*obble*.

spread When you hear a note there is a left-to-right range of where it might be. It seems to occupy all of this horizontal interval. For me C is widest, ear to ear, and F# is narrowest. I am curious whether these particular extremes are reached at the same notes for all people. Call this spread: it is different for different notes, independently of what octave they are in.

Close your eyes and listen through stereo headphones while you hit notes at random on an on-screen keyboard. (It's not hard to make it random: motor control with your eyes closed isn't exact enough to make it anything but, especially if you complicate your mouse movements.) You'll find that the sizes vary between these extremes. Don't worry about making them consistent between trials. That will come, especially given the next factor.

wobble When you hear a note it can move from one location to another, either wandering in space before you or oscillating from side to side. You hear it as being in a certain location and then you swivel your attention to a different direction and you catch it again. Do this with headphones and closed eyes too. For me E rotates through the greatest angle, and B flat through the smallest.

The easiest contrast between spread and wobble is that spread is simultaneous, and occupies the whole angle between its limits. It can move smoothly between them. Wobble on the other hand is alternating. It stops in one location before you hear it in another. These differences may be more obvious at high and low pitches, octaves 1, 2 and 6, 7 on the usual numbering. And they are clearer with artificially shaped waves than they are with real sounds. I recommend listening to pitches as sawtooth and as square waveforms (for example using (ii) or (iii) in appendix II), to make wobble and spread, respectively, vivid. To my ear they are undeniable when heard as these waveforms. The waveform labelled 'fifths', presumably because of the balance of its overtones, also gives salience to W, as sometimes for me a synthesised panpipe sound does. It is significant that they remain vivid when one listens with only one ear, for example by wearing stereo headphones and reducing the volume of one ear to zero. Then you can listen to other pitched sounds, piano, flute, and so on, and hear them as sounding like sawtooth or square waves (listening to those components of the sounds) in order to bring W or S to the fore. (And I find that notes heard with an effort to separate W and S have a special beauty, even when it does not pay off for recognition or memory.)

You may find it confusing when the sound has an inherent oscillation, so that in decaying it seems to be alternately louder and softer. This can sound like wobble. A defence against this is to control the left-right wobble alternation by will, changing its speed or pausing it and the producing a burst of it. (You can do this.) The oscillation of decay is not controllable.

Try this. Play a scale telling yourself that the notes are at first getting 'wider'. Don't define wideness, just listen expecting to find something in the category of size or extent. You should find that the notes get obligingly wider, up to some maximum, and then decline to a minimum, then increase again. The identity of the maximum and minimum should be the same if you go on beyond an octave. Now play the same sequence of notes, telling yourself that they are getting 'thinner'. You'll find that they now get thinner where before they got wider. And the turn-around points, the minimum and maximum, will be near – within approximately a tone – of the previous maximum and minimum. Then try to hold both in your mind at once, best for three notes or so which are uniformly increasing with the one expectation and uniformly decreasing with the other. If you can do this, you'll experience *different* extents increasing and decreasing. So either there are two things going on behind the impressions of width, or the whole business is a matter of suggestion. Evidence against the latter possibility is that you cannot shift the turning points significantly or make arbitrary stretches of increase and decrease, or make different octaves behave differently, or make the impressions differ significantly from one day to another.

In the next section I describe several ways of hearing a note in terms of spread and wobble, in order to classify, remember, and imagine pitches. I am combining a number of separate mental tricks, which have been effective for me. You may get results by combining them in some different pattern. My guess is that if you can make use of my experience, building on the techniques that worked for me and avoiding my many failed experiments and dead ends, you can in say six weeks of frequent practice make the progress that has taken me eight months. In an appendix I list other methods I used at various stages, in case you have to have taken something like the rambling tortuous route I took, in order to get to the same end. The tricks all involve either associating a spatial image with a pitch or linking the perception of music and speech. I suspect that both, the spatial aspect and the phonetic aspect, are needed. In a later section, **11**, I make some suggestions about why they might be. But the first task is to find something that works, rather than to have an explanation of why it might work.

3. a method

The most basic technique is simply to hear pitches, listening for the two components and hearing the notes as having them. You can gain from my efforts here. Comparing notes in pairs and combining the comparisons, and hearing them as other wave-forms, particularly as sawtooth waves, led me to the diagram below, where blue is spread and red is wobble[2]. (The warning in the following section is relevant here. If you find you can separate out two pairs of maximum and minimum separations, but they are quite different from those below. Then my experience will be less of a shortcut for you. See section **4**.)

---

2 There are pairs of S and W that do not correspond to particular pitches. I suppose it is not inconceivable that these might be evoked by suitable combinations of vibrato and timbre, as 'impossible' colours such as bluish-yellow and reddish-green can be produced by just the right combinations of cone-fatigue and image stabilisation. Then our musical experience would be richer than simple acoustics would suggest. They could also be compared to what are called 'imaginary colours', which are regions of colour spaces that correspond to no possible combination of cone activation. One might speculate that these could be more easily imagined than heard, by imposing the required SW pairs on white noise.

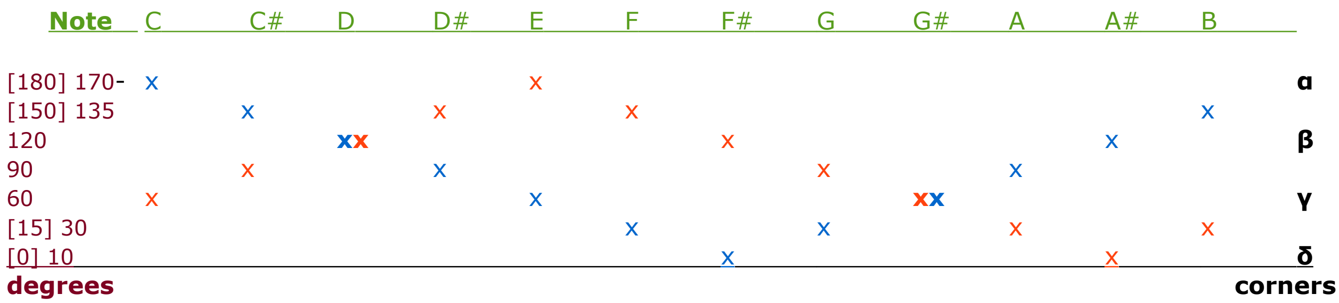


diagram 1

To think of S spans as angles imagine your hands held out at arms' length, hold the span with them, and see how much of a clock face that makes. The Greek letters in the 'corners' column are the measurements of the four extreme values. (Diagram 2 shows why I call them the corners.) They return in section **5**.)

I have drawn the changes of S and W in diagram 1 S as linear, though I am sure they are in fact more sinusoidal, and diagram 2 should surely be more curvy[3]. But that would be theory over-riding the rough observations I want to present. I think the deviations from linearity are fairly small, and confined to the upper end of the curves, because we find it hard to perceive a wobble of a full 180 degrees, and the bottom end, because we would not notice a spread of 0 degrees. (But then sinusoidal patterns will always diverge from linearity more at their extremes.) And these are equally tempered notes: to present the data otherwise would be to claim too much for my capacities. The maximum and minimum of each factor are separated by a half-octave, as should be. Not that the max and min of either corresponds exactly to the minimum and maximum of the other. They are not simple inverses, but the max and min of each is about a tone from the min or max of the other. That is just as well, as otherwise some different pitches would have identical profiles. Diagram 2 makes it clear that no two notes have the same profile.

Instead of drawing curves – and leaving more hostages to fact – I have left the

3 The sensitivity graphs of the three colour receptors, while roughly similar shapes, are more out of phase. Taking the range of perceivable colour as an octave – it does roughly double in frequency – the maximum of the medium wavelength receptor is about a third above that of the short wavelength receptor, and that of the long wavelength receptor is about a fourth above the short wavelength one.

lines straight and modified the scale at the left, indicating in square brackets the values that would follow from full linearity.

Another way of presenting the same information is this.

W								
7			E					
6		F		D#				
5	F#				D			
4		G				C#		
3			G#				C	
2				A		B		
1					A#			
	1	2	3	4	5	6	7	S

diagram 2

Though no two tones have identical profiles, the close spacing of the two curves does make some discriminations tricky, especially between chromatically related pitches, and between pitches which have similar profiles though the absolute sizes are different, such as D and G for me. Notice, though, as a partial mitigation of these difficulties, that most pairs of notes can be distinguished on generally qualitative grounds: E and G#, for example, have about the same spread but the former's wobble is considerably greater than its spread while that of the latter is less and roughly equal to its spread. Note the subtler contrast between D# and F, though. More quantitative considerations seem to be needed there. (There is evidence that some pairs of notes are harder to tell apart than others, for people with AP.)

Practice, noting spread and wobble until you can tell them apart without consulting the diagram. You may not find you can classify notes quickly, but you should find that you can do so consistently. You will make persistent mistakes, particularly in distinguishing between notes in the two half-way stretches between the extremes, and you may find that you hear occasional extremes of one factor as extremes of the other. Consider how often you classify a note as having a profile near to its correct one, though. You can increase your accuracy with practice.

But to increase speed and accuracy it will help to bring in two other factors, one phonetic and one dynamic, in combination. Hear notes as beginning at a definite point in time but having two endings. The note begins with a percussive or consonantal onset sound, which soon ends. Hear the end of this onset as having a sound quality itself. (Remember the sudden palpable silence when a noise that you have become unaware of suddenly stops. Think of the mouth-closing sounds that English speakers put at the end of French final vowels. Think of a glottal stop in an East London accent[4].) You will find that the on-sound and the off-sound occupy different locations. (For me it is a left to right skip.) That is wobble: the separation of beginning of *on* and end of the first *off*. Try encouraging the skip – not a slide – from one position to the other to happen very fast. But simultaneous with this, or beginning a milisecond later[5], there is also a continuous spread, a slide or a spreading, to fill an extent that can end either within or beyond the wobble extent. That is spread.

There is an image that goes with these discriminations. The initial consonantal wobble can present itself as a rapid touching of one point and another, as if two hands were quickly touching the locations. Then the spread is like a glow that appears between or around them, depending on the relative extents, or as a pressure from the two hands inward or outward depending. It can seem as if there are four imaginary hands (!), two razor-thin ones making the W touch in quick succession, and then two appearing near them to press and define S. (See appendix I in this regard, for images I found helpful, and section **11** on phase based location mechanisms.)

4 There is a well-confirmed correlation between AP and speaking a tonal language (Deutsch 2013b). To my ear the contrast between syllables of Mandarin, at any rate, that differ in tone is located at the boundary between the initial consonant and the following vowel, not during the vowel as thinking of them as if sung might suggest. This fits with the 'double consonant end' picture, and suggests that learning such a language a child learns to pay attention to this aspect of a tone.

5 It would make sense that S should appear later than W, on the gloss I gave in introducing them above. If W corresponds to finding alternative directions to search for a sound source, in each of which one then carries out a search of a given spread, then one will often fix W before S.

So: hear notes as having a beginning – an extended percussive/consonantal onset – and two endings. Practice with known notes – the 'four corners' exercise in section **5** is good for this – until you can hear these, and then practice with unknown notes until you can at least approximately classify them. As you do this you can begin to chip away at two further, linked, projects. (a) getting the representation of a note to summon an aural image of the note; mind's eye prompts mind's ear. (b) getting representation and sound to fit, so that you can hear a note as having its stretch and wobble. It *sounds* like this combination, much as a specific shade can come to look like a combination of a particular saturation and a particular brightness, and then that combination presents itself as the way the colour has always looked. I think that paying attention to these further tasks helps with the seemingly less demanding recognition task, as learning the exact attributes of given notes is a good way of fixing them in your mind so you can recognize them. Short term aural memory combines with long term visual memory to foster long term aural memory.

Hear notes as having these characteristics until they begin to sound intrinsic to the sound of the note. Then begin striking notes on a keyboard at random, remaining ignorant of their identities, and classify them, at first with the emphasis on getting roughly accurate W and S rather than naming the note. As the same time spend time with the atonal ear-worms of **5**, trying to fix them in your mind, so that you can then remember them as the W-S combinations that they are.

4. A warning

I do not know whether the extremes are the same for all people, especially given that there are different possible explanations about how the method works (see section **11**, 'why it might be so'.) It may vary with say, the distance between your ears or how strongly one ear, eye or hemisphere is dominant[6]. After all some

6 This is a very basic question. If there are individual differences, it is tempting to look for correlations with those revealed by Deutsch's tritone illusion, which shows systematic individual differences in which of two tones of indeterminate height (Shepard tones), separated by half an octave, is heard as higher (Deutsch 2013b, pp 142-144.) The connection would be via the testable suggestion that people use a heuristic of relying on size of interval to determine which of two

colours look green to some people and blue to others. But it should not be hard to adapt my picture to the way things sound to you. You may find it helps to construct your own diagram analogous to the one above, putting together a number of pair-wise comparisons of neighbouring notes, using sawtooth and square wave generators if that helps, so that the pattern has one maximum and one minimum for each of spread and wobble separated by a half-octave, with the maximum of each being within a tone or so of the minimum of the other.

5. listening and imagining: atonal earworms

Since we often get tunes stuck in our heads and since we seem to remember them with the pitches we have usually heard them to have, you might think that one could learn pitches by remembering the notes of familiar simple tunes. But it doesn't work. (It is an interesting fact that it doesn't work.) Two reasons are the way we adjust tunes to the key of sounds we hear, so though you may hear "three blind mice" in F when all is silent, the moment another tune in another key is heard your grasp of what you were remembering wavers. Combined with this is the difficulty of comparing remembered or imagined notes to those we hear. You hear "mice" in your mind, knowing it is F, then hear a random piano note, and find yourself in puzzled uncertainty how the two are related. (And if you try singing the mice, to compare two physical sounds, you may find that the moment it comes out of your mouth the vivid memory of the tune dissipates.)

I have two suggestions for circumventing this. They have to be combined. The first is to fix in mind short atonal fragments, so that there is no tonic whose shift can throw out the whole process. The other is to learn not just the identities of remembered notes but their W/S structure, and to compare them to heard notes in terms of them.

notes within a half octave is higher, and when they are about equally spaced choose either in terms of S or in terms of W. There would be two determinants of individual differences: where the extrema are and which factor was used. The fine structure of judgements, including where uncertainty is found, would distinguish the use of the different possible heuristics.

In this context a contrast between the methods I am describing and some methods purporting to develop absolute pitch is worth noting. These methods often tell one to listen to a single note and concentrate on it until one can recognize it, and only then move on to another. So all going well one gets single notes one at a time with full accuracy. I have doubts about how well or often this will work, and indeed about the honesty of many such enterprises, but I have not seen quantitative studies of them. As I remarked, my emphasis is on access, letting accuracy come as it will. But it is also global: the conscious judgements of the two qualities come in degrees, and can be refined with respect to a number of notes simultaneously. In fact, it is best done for several at once, as the comparisons of spread and wobble often give information about the character of a note. ("It has quite a wide spread, but not as wide as E, so it is D or F.") Doing it this way allows memory for intervals to scaffold memory for pitches. This is possible because we can use an approximate grasp of any of three factors – W, S, or interval from a previously identified note – to give an approximate grasp of the others, which can then refine the original estimate. On the more standard approach there are no descriptions suitable for approximation, nothing that comes in indefinitely small increments, to refine and set up a mutual nudging[7].

(i) *Preliminary.* Take three notes. (I worked with E, A, C, because of a website with these three as digital tuning forks. See appendix II below. But they are contrasting pitches on my classification.) Play one, hear it in terms of S and W, and combine these until they merge into the way it sounds as a musical note. So you are now holding the sound in your mind but hearing it as consisting of the two components. As it begins to fade in your aural memory use the two components as ways of bringing it back. When you can do this reliably, summoning it when it is no longer in

7 This might explain why some people with absolute pitch find music that is slightly out of tune or not at concert pitch more disturbing than non-AP people with good relative pitch do (Deutsch 2013). The explanation I suggest would trace this to learning a range of discrete notes and not having a good grasp of the continuous variation beneath them. People with a more global grasp of pitch, on the other hand, will be able to hear a note as a certain perhaps approximate degree of spread and wobble, which allows it to be grasped even if it is not one of the perfect exemplars. Of course people with really wonderfully accurate AP will be able to do this too ("that's how A would have sounded in 1860") and there are stories of such people rejoicing in exotic mis-tunings.

consciousness, do the same for another, then the third. Lie in bed willing them into mind, in various orders, to make basic three note tunes. Then when this is solid – the next day, perhaps – go back to the source of the notes, first summon them mentally as vividly as you can, and then see if they fit the notes when you play them. This is only a warm-up, and is likely to collapse in the presence of real heard notes and tunes.

(ii) *The four corners*. The notes with extreme values of W and S are for me C (max S), E (max W), F# (min S), and A# (min W). (If they are different for you, modify this exercise.) The other values of the profiles of these notes are intermediate between the extremes. Thinking of both S and W in terms of angles extended or oscillated through, we have four values. Call them alpha – about 170 degrees or just less than 9 to 3 on a clock-face; beta – 90 degrees or 12 to 3; gamma – 60 degrees or 12 to 2; and delta – 10 degrees or 12 to less than 1. See the remarks in section **3** on measuring S in degrees. As follows:

	C	E	F#	A#
W	Gamma 60	Alpha 170	Beta 120	Delta 10
S	Alpha 170	Gamma 60	Delta 10	Beta 120

It helps to learn the angles as felt rather than as calculated, so that the correlations with these pitches are right. (Especially for alpha and delta where the suggested departures from linearity are surely not precise.) And it is important to remember the notes in terms of the sizes of the angles rather than as remembered tones. For most of us the latter will not work, at any rate until pitch sense and awareness of spread have become thoroughly integrated. Try spreading your hands and arms to give a tactile grasp of the angles. They can then be used as guidelines for the spreads of other notes between them in size. Note that before the corrections for non-linearity at the extremes these descend in 60 degree steps: before correction 180-120-60-0.

Produce these notes and fix the sizes in your mind. (Together, they sound rather Debussy-like.) Again, sawtooth helps with W, and then the angles can be remembered and applied to S. Choose two – say C and E – and hear them as combinations of these precise values of S and W. Do this for 24 hours until you can summon these two notes at will. Be patient. Lying in bed and sitting in the sun, letting it take its time. When you are confident of these add the other two. (I find E the easiest to summon. Then C is E backwards, with W and S reversed. F# straddles the S component of E, and A# is F# backwards. Don't add them too rapidly: each additional note will take at least two days.) Then summon them in arbitrary orders, as little unexciting tunes, E, C, B#, F#; F#, C, A#, E, and so on. Don't move on till you can do this.

A crucial moment in this process comes when the notes that came easily not long ago hide. Perhaps they were vivid and effortless when you went to bed and in the morning when you request their presence they refuse to appear. Don't rush to keyboard, computer, or tuning forks to find them. Instead try this. First remember the desired S and W as purely spatial. Then think of a pitchless sound, the buzz in your head or an imagined hum, or an imagined voice or sound some way distant, and combine it with the desired S and W. Impose the S on it and hear it, having determined in advance the W that is going to apply. Make the sound take that spread, keeping the wobble in the back of your awareness. You will find that the sound assumes a pitch. *Now* go and check whether it is the pitch that you wanted. When you can do this reliably you know you are making progress.

(iii) Now add D and G# to complete the whole tone scale that the four corners are most of. These two notes have in common that their W and S are the same, but for D it is at Beta and for G# it is at Gamma. The extents or angles are as in diagram 4 below.

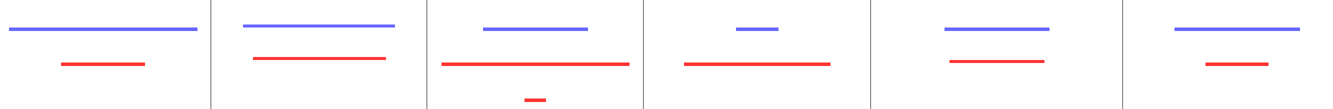
C alphagamma	D betabeta	E gammaalpha	F# deltabeta	G# gammagamma	A# betadelta
					
α β γ δ δ γ β α	α β γ δ δ γ β α	α β γ δ δ γ β α	α β γ δ δ γ β α	α β γ δ δ γ β α	α β γ δ δ γ β α
	both at beta	opposite of C	both shrink by 1	both at gamma	opposite of F#

diagram 3

(iv) Now you can add some familiar tunes, tonal ones, but you have to think of them in W/S terms, and they have to be simple tunes that are vivid for you. And you have to think of them in relation to the five-note scale. (Which is symmetrical in that no note has a unique position in terms of intervals.) You have to start off with the five notes, or some of them, and then switch in imagination to the tune, or part of it, and then back again. I used the following, which are all written out in part (v) of this section. (a) a whole tone sequence in the order F#, A#, G#, D, E, C. (Think of it as: smallest S, smallest W, equal but mid-sized W and S, equal but largish W and S, largest W, largest S.) (b) the last two bars of the Westminster chimes, full hour, in their usual key of E – G#, E, F#, B, B, F#, G#, E. (c) a theme in G from a Beethoven quartet, that has long been vivid for me. And (d) Three Blind Mice in F. These four overlap and come within a semitone of one another at several points, notably several of the four corners, though they are in different keys. You can jump from one to another at several points. To do this you have to think of them as the pitches they are rather than as elements of a tune: you have to detach them from the keys of the tunes. (Just try doing it, and you will see what I mean.) Try to sing them, with the sideways jumps.[8]

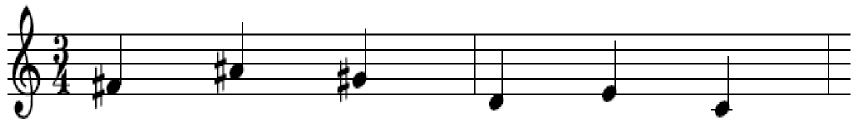
8 The transition from remembering tiny tunes to remembering notes is analogous to the 'holophrastic' hypothesis in Arbib 2012, according to which complex phrases, taken as semantical units, can generate simple words. So what is the analog on that hypothesis of the distracting effect of tonality? Conjecture: the emotive tone or action theme associated with a phrase, where for example fear blocks the extraction of a phoneme signifying 'leopard', applying to both live and dead leopards, from a phrase meaning “leopard nearby, run and climb immediately”.

(iv) This completes my description of the method I have been using for pitch perception. (But see appendix I for other related techniques that I did not stick with. And see section **8** on keys.) At the end of it, I could often identify the keys of remembered tunes and often identify single heard notes. (For heard tonal melodies I try to get the key right, and identify the notes in terms of it. For heard melodies without a clear tonality I am hopeless.) I use a mixture of two methods. (They reinforce one another, if used regularly.) I use the methods of this section to produce mental images of notes – playing an imaginary keyboard – to find the best fit with a heard note. Learning and conceptualising the remembered notes in the way I have described makes this possible, keeping enough comparison and independence between the heard and the remembered. I have to be able to imagine notes a semitone away from my remembered reference notes. And I have to make myself resist the tendency of rhythm and melodic contour to prompt a key and a melody that is often not the one I am trying to bring to mind. (Keeping melodies (b) and (c) separate, going from one to another at will, is a good exercise since they have in part very similar contours.

I estimate the W and S of notes, both heard and imagined, directly, as earlier described, as well as singing and bringing to mind examples like those of this section, thinking of them as seqwuences of particular pitches rather than as tunes. I try to do both and let them support one another. I think that if I continue doing this for years the two will fuse, so that comparison with remembered pitches, with its consequent categorical perception, dominates for quick and regular use, and retreat to the underlying factors is invoked only for slower purposes where finer discriminations are needed.

(v) Four tunes, in different keys but with overlapping notes. The idea is to hear in your mind a few notes of one and then switch to another. Easiest at the point where they have a note in common or a semitone up or down leads to the other. (Beware of learning these mixtures as tunes themselves.)

(a)



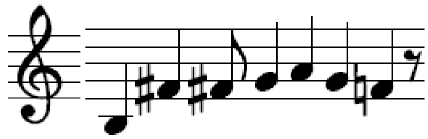
A whole tone tune with the four corners and two W/S balanced notes. Every note occurs in one or more of (b), (c), (d), and is thus an occasion for switching.

(b) Westminster chimes, last two bars of full hour.

(c) from Beethoven quartet op 130

(d) three blind mice, as it comes into my head, with notes in common with (a), (b), (c) highlighted.

(e) tiny fragment made from all of the above, each evoking a different key. You have to keep making new ones, lest you learn them as chromatic tunes in their own right.



6. avoiding mistakes

You will find that you sometimes confuse the extremes of the two kinds, judging wide spread for large wobble, and vice versa. And sometimes you will get the absolute sizes wrong. Here are three ways of minimizing mistakes.

Wait for the pause Wobble alternates and spread runs through the positions between its end points. So make the W jump quick and discontinuous, and make sure that each spread stretch occupies all the space between its end points.

Find the middle A wide wobble can stretch one's judgement of spread and a wide spread can stretch one's judgement of wobble, because one is inconsistently assessing the angle between extreme right and left or between the two inside end-points. The solution is to assess wobble consistently from the mid-points of the squeeze and spread from its end points.

Don't pre-judge it Especially if you are doing too many in a series, you can form judgements based half-consciously on what one takes the note to be. This can result in interval judgements infiltrating pitch judgements. So make practice sessions short, especially at first. (Four minutes is good; but it is hard to stop at four, especially when you are making mistakes and your pride requires that you get enough right.) This third factor combines with a fourth: we get tired. The thinking and imagery here is new, and one is learning something of a kind one has never learned before, so only so much is possible before one starts to deceive oneself about whether one is really following the instructions.

scale Some sounds seem very small. A faint and distant sound at first presents itself as having next to no extent, so nothing to extract S and W from. But once you start thinking in terms of S and W you'll find that nearly all piched sounds, and some unpitched ones, will present the two extents if you ask them to. You have to try it until it works. Then there are really only three scales: the punctuate perceived location, no doubt based on loudness in the two ears and other factors that plan no role in these concerns, a one-sided hemisphere, and a two-sided panorama. Decide on one of the latter two, preferably the last, which is easiest if it is centred straight ahead, and insist on it.

7. Pitch qualities

W and S can be compared in rough as well as precise ways, and as a result we can define a number of attributes of notes that are broader than their exact pitches. Some are uninteresting('somewhere around G') and some are academic ('W exactly 30 degrees greater than S'). I am interested in attributes that have some connection with the sound of a note as the pitch that it is, so that one can apply them by ear, in analogy with rough colour attributes ('blueish', 'pale') which one can apply by observation. Here are my candidates.

Wide or narrow W or S. Thus four attributes. They coincide with 'near E' ('wide W'), 'near C' ('wide S'), 'near F#' ('narrow S') and 'near A#' ('narrow W'). We might call these '**W**ide', '**S**pread', '**S**lim and 'Narro**W'**.

Similar W and S. These coincide with 'near D' (both components fairly wide, and similar), and 'near G#' (both fairly narrow, and similar.) We might call these 'big balanced', and 'small balanced'.

W larger than S. D# to G. 'shrinking'

S larger than W. A to C#. 'growing'

I have had some success applying these independently of pitch. (Not tremendous success.) With shrinking/growing pairs, such as G and A, E and C, F# and A#, the contrast between them seems palpable. In the shrinking member of such a pair the S aspect seems to grow inside the W alternation, constrained by pressure from it, and in the growing member the S aspect seems to fit around the outside of the W. (In this connection see the discussion of tactile aspects in appendix I.)

Note how the whole tone scale that fills in the four corners features explicitly in the first six. And it is indirectly relevant to the last two, in that the boundaries between them are at D and G#.

There is a philosophical aspect to such pitch qualities that is worth mentioning. Aspects of our experience have perceived qualities (green things don't just look like other green things, they look green, and do so even when we know they are not in fact green, as with white walls under green light.) One approach to such qualities, with which I am generally in sympathy, makes the objective properties primary and the perceived qualities only graspable in terms of them. We understand green experience as the way grass and emeralds look (Harman 1990, Martin 2002.) But the attributes I am discussing produce worries for this attitude. It can be illustrated with W and S. Sounds are not really wide or narrow, and do not really jump or spread. If we take our impressions of pitch to be a compound thing, with these as its components, they are formed from parts that are purely phenomenal, which cannot easily be described as "the way things sound when they ...". (Pitch, on the other hand, is the way things sound when they vibrate at various frequencies.) I make some remarks to lessen this discrepancy in **11**.

8. keys and harmony

Distinguishing between W and S helps identify keys too. The method is the same, though slower. I encourage music to jump around horizontally, imagining a location focus that appears and disappears at particular locations, and then when this falls into a pattern I imagine a vertical line moving smoothly on the same axis, centered at the middle of the jump, until that becomes stable. The first is the W value of the tonic and the second is the S. This often gives roughly accurate results, both for identifying the keys of simple tunes and for noting the modulations in more complex music. I also sometimes imagine a cadence and apply the mental keyboard technique to the imagined tonic.

There is a natural explanation of why this should work, for straightforwardly tonal music. It's Pythagorean: harmonies as ratios of small integers. The majority of notes are invariably not the tonic, unless it is extraordinarily boring music. But many notes are related as fourths and fifths to it, or have overtones that are. So if we suppose that the W and S intervals of notes related as fourths and fifths are in ratios of 4/3 and 3/2 then when one slides them around in their apparent locations they should reinforce one another on the W and S locations of the tonic. Diagram 1 above suggests this, but is not precise enough to demonstrate it. (Even with equally tempered notes F4/C4 for example is 1.33 and G4/C4 is 1.49. Pretty near the Pythagorean ratios of 4/3 and 3/2)

Fourths and fifths sound like this to me. (For what that is worth.) When I play a fifth chord the W and S of the component notes with the smaller spread sound as if they are half way along those of the larger one. And for a fourth chord they sound a third of the way along.

I suspect we process small intervals – tones and semitones – differently from large basic intervals – fourths, fifths, octaves, perhaps thirds. The latter are innate in the Pythagorean way described, and the former are learned as one becomes familiar with the scales of ones culture. One can certainly learn to sing and imagine semitones, in order to make the jumps from one key to another described at the end of section **5**.

9. Singing

It is easy to deceive yourself that the note you have just imagined is the note you are now playing, rather than your imagination having shifted to match the sound. The obvious defence is to sing the note you imagine and compare it with the note you hear. I have never been good at singing remembered and imagined notes. (Strange and disconcerting moments when I ask someone to remind me of the identity of some famous tune but they do not recognize it from my attempts.) So learning to let the S and W characterization of a note guide its production is not an idle extra. This is still the aspect that I have made least progress with, and it still is an effort for example to sing two of the earworms of **5** with similar pitch contours without the other emerging, even though they were distinct before I opened my mouth. But with effort I can do it.

I find that thinking of the two component description while reproducing a note helps. Since beginning to do it I am rather less hopeless. The core, I'm sure, is just to keep the two separate and in mind while producing the note, trying to match them and not worrying about making a nice sound or doing anything closely resembling singing as one knows it from good singers. (Perhaps the aim of singing like these people gets in the way. In fact the sounds that I find most easy to match notes with seem to me like ethnomusicological examples of vocalisation from cultures not focused on our vowel-centered ideas of tuneful singing.) Then eventually one will train oneself to do it. There must be many distinct elements of vocal control involved, and one can't expect to be aware of many of them.

For me, it helps to keep in mind the double-ending representations of section **5**. Think of an extended moving consonant. (See speculative footnote 4 on tonal languages.) I find this works best if the noise I produce is a whisper, a hum, or faintly sung speech, nothing like singing to an audience[9]. So my advice is: don't try to sound beautiful, think of the W/S character of the tonic (for that kind of music), and don't neglect the beginning of the sound.

10. Imposing the patterns

Because this structure gives a single image rather than a pair, it can be superimposed on a variety of sounds. In particular, it combines nicely with white noise. Do this: turn on a white noise generator. (There are several on the internet: I find the variant called 'brown noise' works best for this purpose. But an electric fan will do.) With that in the background produce some clear pitches. Fix in your mind the spread and wobble of the pitches. Then superimpose this shape on the white noise: hear the attributes onto the noise to make it take the pitch. Hold the pitched noise after the clear pitch has faded, and then bring it back again. Do this with sequences of pitches. (Best if they do not have a structure of easy intervals, to prevent yourself cheating, as for example (a) of section **5** (v) and variations on it.) What I find is that doing this produces vivid echoes of the original pitched notes, with their original timbre, and when this fades the pitches can still be recovered from white noise considerably later by thinking of the intended structure as a way of summoning the exact pitch-shape. The aim is to expand the number of notes that you can summon, until you have a mental piano you can play at will[10].

The experience of hearing these ghost pitches emerging from white noise is quite different from that of remembering a melody. It is more 'external', less memory-like (as if there is a phenomenology of the reconstructive aspect of memory) and more neutral as to timbre. You can superimpose it on un-pitched timbres, in particular on the high white noise of blood circulating in one's head that many people can hear. (I

9 There *might* be a connection with the well-known phenomenon that people often sing well-known songs in the keys of their famous recordings, though they could not name the keys. The connection would be that people have motor representations of what it would take to reproduce the song.

10 And then comparing heard notes to its output, classifying them by 'analysis by synthesis' (Bever & Poeppel 2010.)

nearly always can, and I suspect everyone can if the background noise is subdued enough) So one way of summoning a pitch at will is to impose the relevant shape on this home-made white noise. (I would not recommend concentrating on summoning any single note in this way, rather than the skill of summoning one from a variety, for fear of inducing tinnitus. You might turn into Robert Schumann[11].)

You can progress from loud white noise, to softer white noise, to pitch-neutral sounds around you. And eventually to the gray screen of your imagination. It takes a lot of repetition. An important step occurs when you manage to hear the two attributes as part of the phenomenal quality of notes. This takes a certain amount of persistence and experimentation. I think you have to stumble for yourself on whatever mental trick is needed. (For me, the trick has something to do with thinking of wobble first and spread second, and hearing the transition between the two as producing what the note sounds like.)

I find that although I am good at distinguishing musical instruments by timbre when I hear them, and as an ex-oboist (never a very good one) I can sometimes distinguish national styles and schools of oboe tone, I find it considerably more difficult to *imagine* particular instruments with any vividness. For me, various un-pitched noises are easier to imagine. A list of typical noises is:

a wooden ball dropping onto a wooden floor
a small metal object dropping onto a hard floor, such as a stone one
a pebble dropping onto a stone
water dripping into water
impact of metal on hollow metal poles, e.g. scaffolding poles
a knock on a wooden door
footsteps on gravel

11 Remarkably many composers have developed tinnitus. A quick search produces Beethoven, Janacek, Boyce, Holzbauer, Draessiker, Fauré, Smetana. I conjecture that a common factor is excessive attention to the sounds in ones head.

some unbeautiful bird cries, such as seagulls and crows
-- and nearer to determinate pitches --
wind noises, including the sound of blowing over the top of a bottle
a wineglass being rubbed around its rim

I expect everyone can produce their own list. I can imagine a series of pitches from white noise, and then superimpose them on sounds like these, to produce 'instrumental' versions of the sounds. One can see how with a lot of practice this could develop into an orchestrator's imagination.

All my life I have felt that some scenes look like sounds. These are usually scenes with some periodic scattered aspect. Moving highlights on water, light filtering through a canopy of leaves, moving shadows, vertical blinds moving in the wind. (There is a Japanese word *komorebi*, meaning 'The scattered, dappled light effect that happens when sunlight shines in through trees'.) Since beginning to hear spread and wobble, more sights prompt a sense of imagined sound, mostly objects moving against a background (people in a park), bright three dimensional arrays in a dark space (arrays of stars seen so that some stand out as nearer and some as further away), and gaps between approaching and receding objects. (Driving towards trees, seeing their branches as separated in receding distance and these separated gaps as moving towards you.) All of these can be seen as evoking my two qualities, as moving spatial processes. (I think motion is essential. Hearing and sight are related in different ways to time. Everyone knows this and it is part of the difference between music and painting, or even music and silent film. But it is hard to state in a way that is more than hand-waving.) But to my initial surprise not many of them gave any sort of representation of pitches. I now suspect this is because pitch is only the surface of a much larger phenomenon.

11 why it might be so

The method I described in section **3** stirs two non-musical dimensions into the hearing of tones. The first is spatial, and the other is phonetic. W and S are described spatially as ways in which the tone locates itself before one. I don't have a better way of describing them, as that is the way they present themselves to me. It is possible though that there is nothing deeply spatial here, and because of some accident of human psychology focusing on the distribution of imagined locations allows one to listen for other not essentially spatial properties of tones.

I suspect the connection is deeper, though. I first began exploring the line of thought that has led to this essay on reading about phase-based location in birds. This builds on the fact that a pure note of a given pitch will interfere constructively with itself at one ear and destructively at the other, if it is at a series of locations with respect to the hearer. (It depends on the pitch and the separation of the ears.) So information about which directions of one's head make a sound loudest generates information about possible locations of the source. For pure notes there is a multiplicity of locations, but this is much reduced for a complex sound, since there will only be a few locations which occur in the possible locations of all its pure components[12]. There is good evidence that birds, especially owls, locate sound sources in this way (Saberi and others 1999, Wagner and Frost 1993.)

If a mechanism along these lines plays a role in human hearing, then the connection between pitch and spatial location would be much less arbitrary. (There is evidence that amusia is associated with deficits in spatial processing. Douglas & Bilkey 2007.) But this particular connection? I am not going to work this out in more detail. It would be speculative, and it would distract from the point that any such explanation could be wrong while the technique of conceptualizing pitches in

12 It is suggestive that small vulnerable birds, that want their presence but not their exact location noticed, tend to call in pure notes, while larger and tougher birds tend to have much rougher calls. No more than suggestive, though: for one thing this impression does not take account of the overtones that are too high for us to hear.

terms of spread and wobble works, allows us to make discriminations and educate our imaginations in ways that we could not previously.

The easy application of pitch information to white noise has a very plausible outline explanation from this point of view. Given a scene in which there are many simultaneous pitches and timbres – a movie jungle soundtrack – a person or other animal will want to be able to locate a particular source in the array. The scene provides the white noise and a briefly noticed or remembered sound provides the source that one needs to separate from the background. So one needs a way of operating on an undifferentiated jumble of sounds to make candidate locations for a target sound salient. Then one can look or reach to that location. And if one misses one needs to a fall-back alternative location.

The other non-musical dimension is phonetic. In section **3** I described W as consonantal and S as vowel-like. And I suggested – see footnote 4 – that hearing a tone as if it had a complex consonantal onset might be easier for speakers of a tonal language. That was speculative, but there is a definite vague resemblance between separating W and S and hearing consonants and vowels as separate components of syllables, which does not come easily to many children learning to read. My own experience of massaging my pitch perception is that combining the spatial and phonetic aspects is essential, whatever its explanation.

There are also much weaker possible explanations. One is simply that there is a systematic structure of contrasts between pitches, and if we fasten on contrasts which are structurally parallel then any representation of the one gives the possibility of transfer to the other (Rosenthal, forthcoming). A related weak explanation, a neurological version of the same, would be that the comparison of periodic information through different channels might be a fundamental neural process, occurring in many different contexts. Then there would be similar neural processing of information in different modalities, which could allow one to be compared profitably to another, if analogies at the right levels are created.

In **7** I described a problem that the pitch qualities I am interested in make for an attractive account of the metaphysics of perception. They seem to be purely aspects of the way pitches appear to us, and not properties of their sources. Phase-based sound location would give some degree of resolution of the problem, if found in humans and if it is an origin of the phenomena I am discussing – note the big ifs. In some circumstances W would actually specify where one ought to look to find a source, thus its range of locations, and S how wide a search in that direction one should carry out, and thus the uncertainty of the location. The 'phenomenal qualities as objective properties' point of view would be presented with less of a challenge, even though in many cases in modern human life that is not the primary use of the information. It is worth pointing out that phonetic qualities in general present similar problems. In contemporary psycholinguistics an r sound, for example, is not thought of as indicating a specific manner of articulation or a specific acoustic property, but as an aspect of the way the hearer classifies the sound. There are connections with articulation and acoustics, of course, but they are indirect. or a discussion of which properties of acoustic signals do represent aspects of the environment,and an interesting contrast with vision, see Bregman 1990 pp 36-38.)

Here is a point of view from which the association between hearing and phenomenal concepts with indirect connections to objective properties should not be surprising. The frequencies of most perceived sounds are well above the repeat rates of neurons. So coding is necessary from an early stage. A sensation of say A4 does not result from anything in the brain oscillating at 440 cycles per second. The sensation is the result of many stages of processing of a signal in which a sound of that frequency is successively coded until it has structural properties that allow it to represent that frequency without having much in common with it. In particular, it is plausible that the process involves sampling from the acoustic wave, by both mechanical means in the inner ear and later by neural means. (Trainor 2015.) I suspect several different sampling processes occur, which makes distinct W and S components less surprising. It also gives ways in which different waveforms, such as the sawtooth and square waves referred to in section **3** could trigger different processes. (Coding to representations that have no intrinsic representation to the properties they represent is true of all perception, but sampling may be peculiar to hearing, both of music and of speech. The frequencies of light are too high to sample from.) So the stages of processing can generate their own phenomenal qualities, and there is no reason to expect that they individually have representational functions. Their primary function is to come together in an eventual representation. Most such properties will inevitably be unconscious, but there is no reason to expect that those that we can access are representational.

(There are three appendices after the bibliography, on initially promising methods I used but did not stick with, on relevant technology, and on a peculiar experience.)

Bibliography

Ali, M, Klyne, M (1985). *Vision in Vertebrates*. New York: Plenum Press

Arbib, Michael A (2012) *How the Brain got Language*. Oxford University Press.

Bever, T and D Poeppel (2010) Analysis by Synthesis. *Biolinguistics* 4, 174-200

Bregman, Albert S (1990) *Auditory Scene Analysis* MIT Press

Davidoff, Jules B. and Fagot, J. (2010) Cross-species Assessment of the Linguistic Origins of Colour Categories. *Comparative Cognition and Behavior Reviews*, 5, pp. 100-116

Dennett, D.C. (1991) Heterophenomenology. in *Consciousness Explained*, Penguin Press.

Deutsch, Diana, ed. (2013) *The Psychology of Music, 3d ed.* Academic Press.

Deutsch, Diana (2013b) Absolute Pitch. in Deutsch 2013, 141-182

Deutsch, Diana (2013c) Grouping mechanisms in music. in Deutsch 2013, 183-248

Douglas, K. M., Bilkey, D.K.(2007) Amusia is Associated with Deficits in Spatial Processing. *Nature Neuroscience* 10, 915-921

Gerstner and others (1996) A neuronal learning rule for sub-millisecond temporal coding. *Nature letters 383,* 76-78

Harman, Gilbert (1990) The Intrinsic Quality of Experience. in J. Tomberlin ed. Philosophical Perspectives, 31–52, 4, reprinted in Ned Block, Owen Flanagan and Guven Guzeldere eds. *The Nature of Consciousness*, 663–76, Cambridge, Mass.: MIT Press 1997.

Hofman and others (1998) Relearning sound localization with new ears. *Nature Neuroscience* 1.5 (1998), 417-421.

Kristan, William B. (1998) He's got rhythm: single neurons signal timing on a scale of seconds. *Nature Neuroscience* 1.8

Martin, M G F (2000) The Transparency of Experience. *Mind and Language* 17, 376–425.

Majid & Burenhult (2014) Odors are expressible in language, as long as you have the right language. *Cognition*

Rosenthal, David M (forthcoming) Quality Spaces and Sensory Modalities. In *The Nature of Phenomenal Qualities: Sense, Perception, and Consciousness*, ed. Paul Coates and Sam Coleman, Oxford: Oxford University Press.

Ross, Alex (2007) *The Rest is Noise* 1st edition , Farrar Straus and Giroux. (The pagination seems to be different in other editions: the quotation at the head of this essay is on p 168 of this edition. Ross gives M Huttunen in *The Cambridge Companion to Sibelius* as his source, which I haven't checked.)

Saberi and others (1999) Neural bases of an auditory illusion and its elimination in owls. *Nature Neuroscience* 2, 656-659

Trainor, Laurel J. (2015) The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Phil. Trans. R. Soc. B* 370, 1664

Wagner and Frost (1993) Disparity-sensitive cells in the owl have a characteristic disparity. *Nature* 364, 796-798

Websites on stereograms: http://www.eyetricks.com/3dstereo.htm
http://en.wikipedia.org/wiki/Autostereogram

appendix I: other methods

I have described the methods for determining W and S that I use now. On the way to them I experimented with a number of variants, which I did not stick with. It is possible that some of these were necessary for me to learn to use the eventual methods. So I describe two common themes of some of them below – touch and three-dimensionality – in case they are useful or interesting.

touch I often found myself thinking of squeezing and spreading the sound. An element of this remains in the method above, in describing W as a rapid alternation in which two locations are touched in rapid succession, and in the suggestion that separations be learned as separations of outstretched arms. I also thought of S as a kind of squeeze, best described as the experience of bringing like poles of two magnets together in which the space between the magnets feels as if it is filled with a springy force. (Aristotle thought that touch was the basic sense – *De Anima* II, 2 – and whatever we make of that thought, touch does connect perception and action fairly intimately, since you cannot separate feeling objects from pressing against them. And unconscious perception can be correlated with actions that we can then be conscious of. Some instrumentalists find their fingers playing notes as they hear them, even though they cannot name the notes without noting their finger positions.)

three dimensions Different 3-D images seemed helpful at various stages, and though none proved really effective in summoning or recognizing pitches, each when it first came to me seemed really to capture a sound-space connection. The remnant in the method as presented here is the touching of locations and the 'four hands' image at the end of **3**. It is natural to suppose that analogues to three-dimensional vision should appear on my approach, since one is comparing the outputs from two channels, giving an analogy with binocular vision, and one is making an unconscious representation accessible to verbal consciousness, for which learned motor routines, which typically occur in space, are a useful conduit. (You often know what you think by noting what you are saying; you often know something is hurting because your posture has changed; singers and instrumentalists can reproduce notes they cannot name by ear alone.) W still seems to me to be a side to side alternation – left to right in fact – in a plane at right angles to my direction of attention, while S is more like the extent of a solid meeting that plane but moving towards or away from me. And when I ignore my methods and theory, close my eyes and ask what a heard or remembered pitch looks like, it always seems like a shape moving in three dimensions. (But different shapes moving in different ways on different occasions!) Also, the 3-D stereograms that were a fad a few years ago just strike me as rather musical. So while three dimensional representations do not play a large role in the method as I am presenting it here they were often in my mind and may be useful to you. Any combination of real or imagined actions, locations and shapes in space, muscular association, and generalization from phonetics is worth experimenting with.

appendix II: actual and possible technology

To practice these things you need to be able to produce named notes and listen carefully to them. You will need to test yourself see how near to getting them right you are. I have found these useful:

i) Online tuning fork: http://www.onlinetuningfork.com/ E, A, C

ii) "Absolute Pitch" app for android. (Don't be put off by the title; that's just to make sales.) I have produced random notes by closing my eyes and sliding the mouse up and down the keyboard. Imperfect motor control and friction conspire to make the eventual note unpredictable. I like the variety of timbres available.

ii) Virtual keyboard:
http://www.bgfl.org/bgfl/custom/resources_ftp/client_ftp/ks2/music/piano/
There are grander piano-like internet keyboards, but I like the variety of timbres available on this, and the chord mode.

iii) an online waveform generator:
http://www.wavtones.com/functiongenerator.php

You can do it all with these. Or others. But there is room for software targeted specifically at the method I am describing. A simple useful program would present choices of notes to distinguish, increasingly many as one had mastered smaller sets. For example one could progress from the four corners of **3** and **5** to the corresponding whole tone scale, and eventually to all twelve. This would be best with tones of indeterminate height (Shepherd tones) to prevent one learning the notes melodically.

I would be interested to discover the uses of what one might call a well-tempered white noise machine: an ever-changing combination of notes from across the concert pitch chromatic range. (Like banging an unevenly shaped stick rapidly and repeatedly against the strings of a piano: you won't get anyone to volunteer their Steinway for this. I had to make do with the chord mode of the virtual keyboard mentioned above.) This would be useful for training oneself to hear pitches on demand. One would hold the specifications of a note in mind and then let it appear from the noise. This can be done with regular white noise, but then there are many more unwanted targets that one has to hear past.

There would be uses for wave-form generators that do what I have used sawtooth and square waves for, to make W and S evident. There might be wave-forms that made each of these even more evident. There might be wave-forms that made both simultaneously evident and different. Then one would progress by little increments to more musical-sounding notes. This would combine naturally with the 'equally tempered white noise machine' described above.

A radical idea would be a sound generator that presented a particular spread or wobble. It would give a white noise-like array of pitches – best along the lines of the well-tempered machine described above – which all had in common a particular S or W. Then one could hear these in isolation from hearing a pitch. They would be like Shepherd tones (Deutsch 2013c), but more specialised. One could test oneself: what is the S or W here? And one could demonstrate the components: S is what all these have in common, and W is what all these others do.

appendix III: an auditory-tactile-visual experience When I was working on 5 I made a preliminary version of diagram 2, to fix in my mind the sequence of separations in the whole tone scale that grows from the four corners. I rehearsed using the diagram to summon the tones, as described, for an evening and went to bed. Then for two hours I felt as if I was seeing and touching the notes, as if they were vividly and tangibly before me as large spatial shapes moving around me. The shapes were silent, but attending them made it easy to summon the notes in my imagination. In spite of the vividness of the shapes I could not determine which were their relevant geometrical features. This continued in hallucinatory fashion till I finally slept.