Individual differences in pitch perception, and their possible relevance for language typology

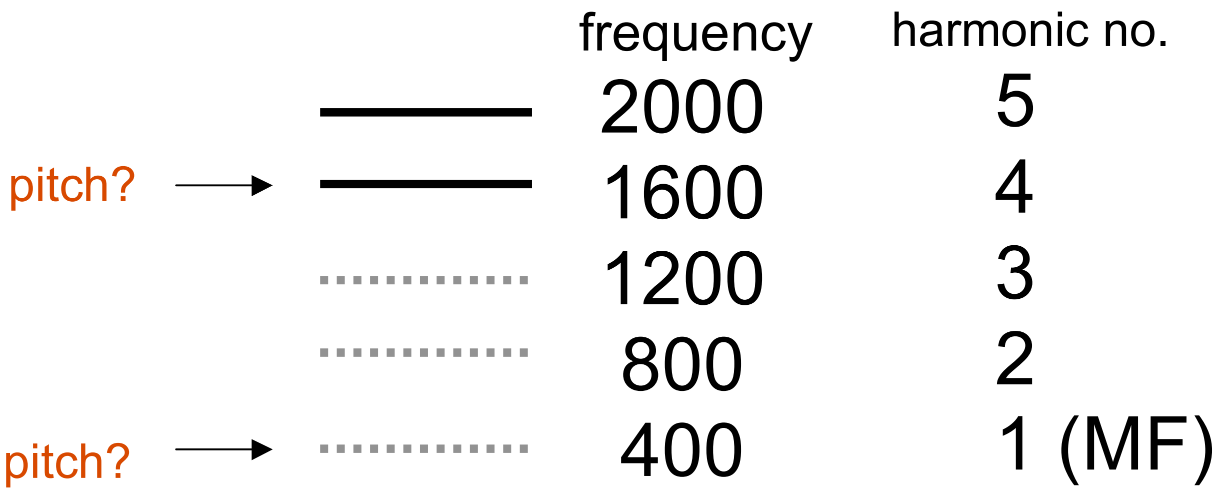
Bob Ladd, University of Edinburgh

Transcript of talk presented at the 27th annual meeting of the Phonetic Society of Japan, Kanazawa, September 2013; please do not quote verbatim.

The topic I want to discuss today revolves around the question of whether individual differences in perception and cognition might be relevant to the design of language. Specifically, I want to discuss some recent work on ***stable individual differences in pitch perception*** that might be relevant to the way pitch is used in language.

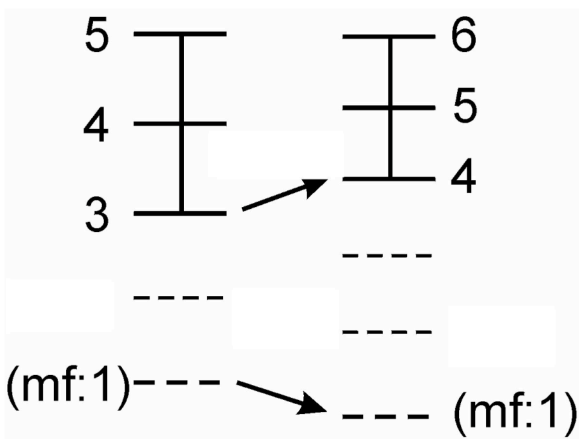
In the first half of my talk I'm going to present recent findings on the perception of ***pitch*** in stimuli that physically don't have a ***fundamental frequency.*** In the second half, I'll discuss some ideas about possible links between population genetics and language typology, and then describe how the pitch perception work is linked to it.

Now as I'm sure you all know, in general, the percept of pitch is linked to the fundamental frequency of an acoustic signal. But it's possible to construct stimuli that in some important sense don't have a fundamental frequency. These ***missing fundamental*** stimuli are tones that consist of a small number of frequency components that could be multiples of a fundamental frequency, but where that fundamental frequency is not one of the components of the stimulus. This is shown in the following diagram:



When presented with a stimulus like this, some people will perceive the pitch of the stimulus to be that of the missing fundamental frequency, and others will perceive it to be one of the components actually present. Here are some brief references to this work – there's more detail on your handout. (refs: **Smoorenburg, Schneider, Seither-Preisler)/**

Smoorenburg devised a simple way to test people's percepts in his study in *JASA* in 1970. What you do is give people a sequence of ***two*** missing-fundamental tones, constructed so that if people are hearing the fundamental they will say that the pitch goes down from one tone to the other but if they are hearing the component frequencies actually present they will say that the pitch goes up. So for example:

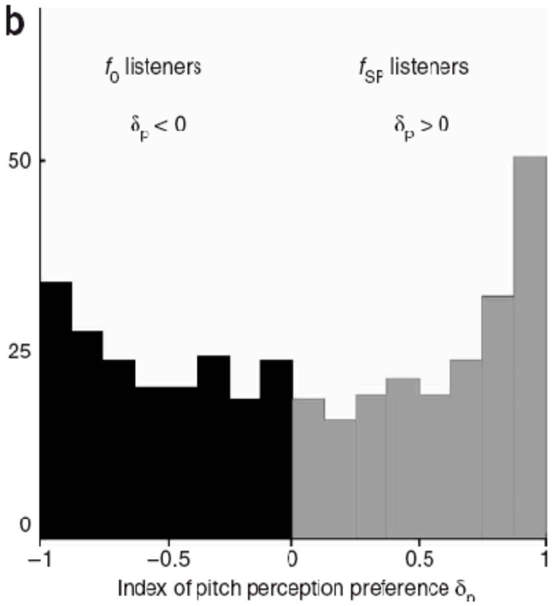


Here's what these sound like.

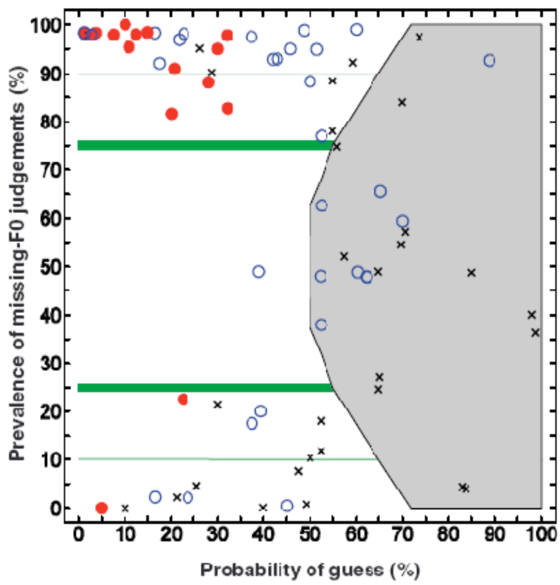
AUDIO DEMO (stimuli available at http://www.lel.ed.ac.uk/~bob/JEP/)

So you can see from looking around that not everyone hears these the same way!

Several studies show that these differences are fairly consistent and reliable. Schneider et al. refers to people as "F0 listeners" if they regularly hear the missing fundamental and as "spectral listeners" if they normally hear one of the component frequencies. I'll use this same terminology here. Also, Schneider's group came up with a way of quantifying individual behaviour, which assigns everyone an index ranging from -1 to +1, and I'll refer to this as the **Schneider index**. An index of -1 means you always hear the missing fundamental as the pitch of the stimulus, and an index of +1 means you never do.



Schneider et al. 2005



Seither-Preisler et al. 2007

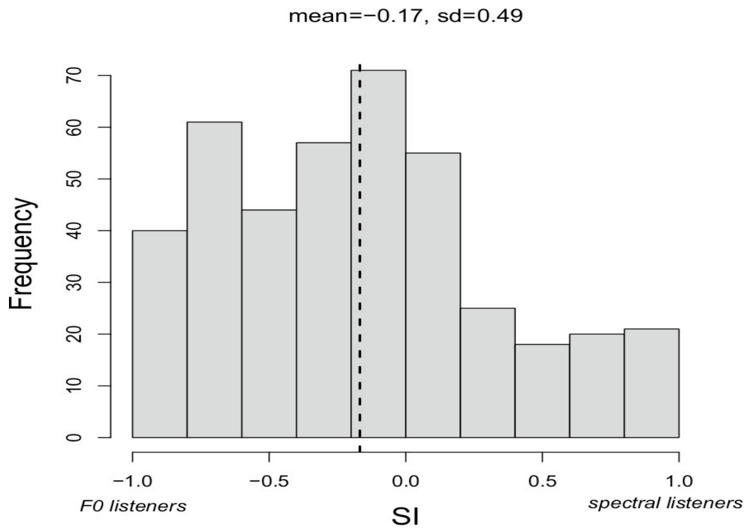
Now, an index of zero means you don't have a consistent preference. Actually, that creates some interesting problems of interpretation. Both Schneider et al. and Seither-Preisler et al. report bimodal distributions, with more F0 listeners and spectral listeners and fewer people in the middle. But in quite different ways they excluded some of those 50-50 people from their analyses, and as we'll see shortly this may throw away some interesting phenomena.

Anyway, after I read about this work, I started experimenting with stimuli like these myself, in collaboration with Dan Dediu. With several other colleagues, we've now run several experiments similar to Schneider's and Seither-Preisler's, with a total of over 400 listeners. Our paper on this work appeared online in *JEP:HPP* in February. Details are on the handout.

The main thing we did differently from Schneider and Seither-Preisler was to have only a small number of different stimulus types – 10 or 15 – and to have listeners judge the same stimulus a total of ten times. This gave us more insight into the behaviour of inconsistent listeners, ones with a Schneider Index around zero.

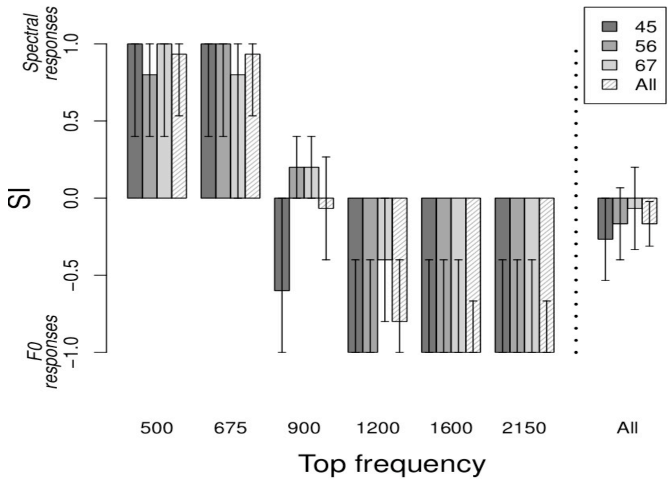
Our basic findings are as follows. First, there really are people who are consistently F0 listeners or spectral listeners. Here's a histogram of the Schneider index from the more than 400 listeners in our data. But you can see that there are quite a lot of people whose responses are mixed in various ways, which gives rise to Schneider Index values in the middle of the distribution, closer to zero. We definitely haven't got a normal distribution, but it's not really bimodal either. As I said, some of these people in the middle are excluded from both Schneider et al.'s data and

Seither-Preisler et al.'s – in fact, Seither-Preisler et al. excluded about 25% of their subjects on the grounds that they're just guessing. But our results show clearly that patterns of mixed responses are quite systematic. In other words, it's not just a matter of guessing or simple inability to do the task.

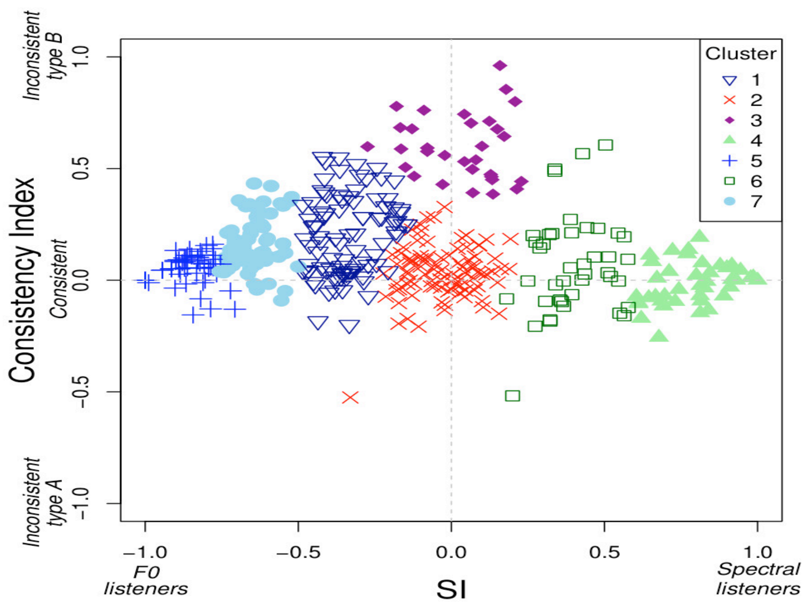


Among other things, we've found that, regardless of their listening strategy, people's Schneider Index is very stable. We did an analysis of test-retest reliability in one of our experiments a found a correlation of .87. Schneider et al. also report high test-retest reliability.

Because we didn't exclude any subjects, we've been able to identify some of the factors that are relevant to understanding the patterns that give rise to a mid-range Schneider Index. One common pattern of mixed responses is to give spectral responses when the frequencies in the stimuli are mostly below about 1000 Hz, but to switch over to F0 listening with stimuli containing higher frequencies, as seen the following graph of results for one individual subject:



We've quantified this tendency to switch response pattern as the frequency goes up with another index that ranges from -1 to +1, which we call the consistency index. The consistency index is also highly reliable – it gives a test-retest correlation of .94. We've also been able to show that the Schneider Index and the Consistency Index account for much of the variance in the data, and that they're the main dimensions that influence a cluster analysis of our subjects. This means we can usefully place individuals in a two-dimensional plot showing their overall Schneider Index and their Consistency Index.



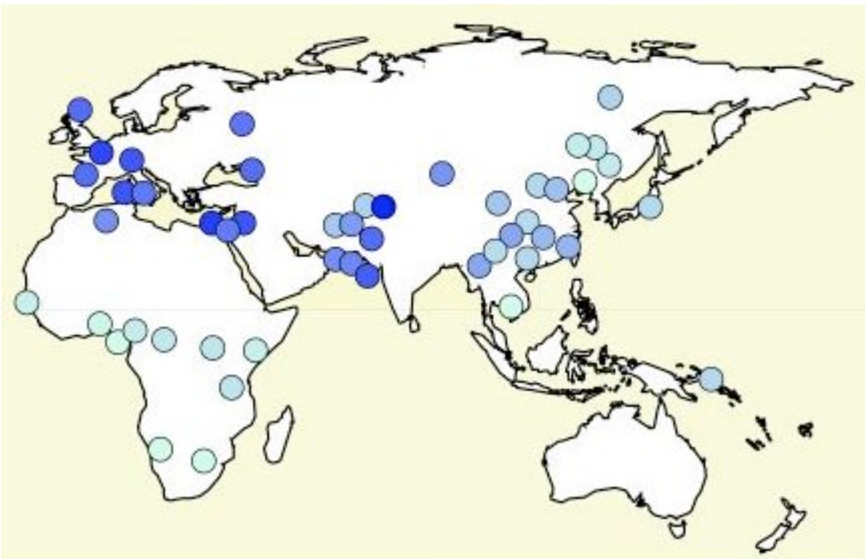
So we've still got people who don't show a consistent pattern of preferences. There must be other factors as well that we still don't really understand, and clearly, the Schneider index by itself is not an adequate characterisation of the range of individual differences.

There is much more that I could say about our experiments, but for the second part of the talk I want to consider the relevance of our experiments for language and language typology. The key feature of the missing-fundamental task is that it involves some sort of interplay between determining the ***spectral properties*** of an acoustic signal and determining the ***pitch***. We think that this is relevant to normal speech perception. If you think about it, any time we perceive the multiple differences in a pair like /â/ and /ĭ/, we are simultaneously making judgements both on the basis of the spectral properties of the two vowels – that's the difference between **[monotone]** /a/ and /i/ – and on the basis of the fundamental frequency – that's the difference between **[hum]** \ and /. What the missing fundamental task shows is that individuals may have slightly different ways of going about making such judgements – in other words, in some way that's still far from clear, the balance between the spectrum and the fundamental periodicity is computed or processed somewhat differently by different people.

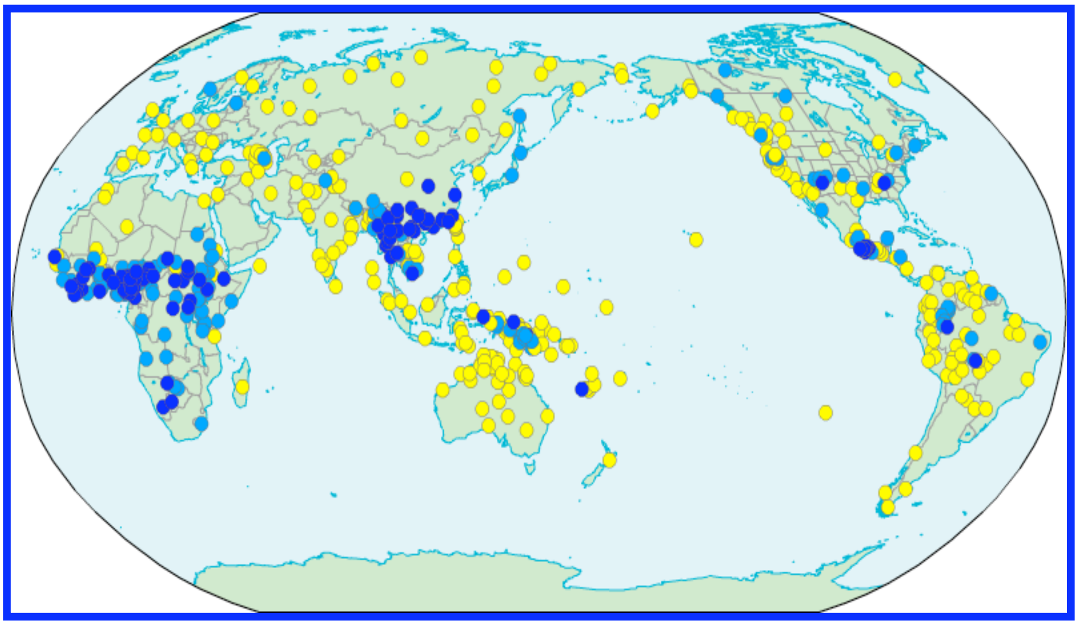
And this is what leads into the topic of language typology. As some of you may know, about 6 years ago Dan Dediu and I published a paper about a possible link between population genetics and language typology – specifically, a link between the geographical distribution of tone languages and the distribution of older and newer variants of the genes *ASPM* and *Microcephalin*, which are involved in brain development.

This paper had its origins in 2005, when I saw a report on work by Bruce Lahn's genetics group in Chicago. Lahn's group showed that genetic variants of these two genes are very unevenly distributed in the world's populations.

Without going into a lot of detail, I can summarise the Lahn group's findings as follows: There is an older and a newer version of both of these genes, with the newer ones apparently having evolved within the last 20 or 30 thousand years – probably rather more recently in the case of *ASPM*. The newer versions are most prevalent in Europe and Northern and Western Asia; the older ones in Africa and SE Asia. (Lahn's group had very limited data on New World peoples and I'll only talk about the Old World distribution.) Here's what their population genetic data for *ASPM* looks like on a map:



Now, when I saw these maps, I thought: *These maps look a lot like maps of the distribution of tone languages.* In particular, the older version of the genes is prevalent in places where there are tone languages, and the newer version in places where there are non-tone languages (tone language map from WALS reproduced below). Could it be that, in some way, population genetics is related to the way human language is manifested in different populations? And if so, what kind of influence could a gene have that would induce people to treat pitch differently?



This is the question that Dediu and I investigated. Using a lot of data on the genetic makeup of the populations studied by the Lahn group, and a lot of structural data that we gathered on the languages spoken by those populations, we showed statistically that there really is a correlation between the geographical distribution of the variants of *ASPM* and *Microcephalin-1* and the geographical distribution of tone languages – an association that is unlikely to be due to chance.

Well, OK, but the obvious next question is: what kind of influence could a gene have that would induce people to treat pitch differently? In our paper Dediu and I acknowledged that we didn't have an obvious answer. What we emphasised was that, whatever the effect is, it would manifest itself only through the process of intergenerational transmission, favouring particular directions for language change.

This is an important point that some of our critics didn't really take on board, but I don't think it should be especially controversial. Obviously, any normal human being acquires the language of the community they grow up in, but at the same time it's widely accepted that at least some language change occurs as a by-product of the acquisition process. We suggested that the genetic differences might underlie what we called a "cognitive bias", which would incline people to treat pitch differently, and that this might influence language change over historical time.

But even if this is the right way to think about the problem, we really had no basis for saying what the cognitive bias might be. All we had done was propose that population biases ***could*** influence language change, and show that the geographical association between genes and tone languages was unlikely to be coincidence. The ***nature*** of the bias remained mysterious, and the actual connection from genes to bias was pure speculation. And this is why we were interested in the reports by Schneider et al. and Seither-Preisler et al. that there seem to be different ways of perceiving missing-fundamental stimuli. Those differences could point to the nature of the hypothesised bias.

+++

Things have moved on a bit since then. First of all, Patrick Wong and his colleagues recently reported finding a link between *ASPM* and pitch perception. Their study was directly motivated by the Dediu and Ladd paper. Very briefly, Wong et al. provided experimental evidence that there are individual differences in the way pitch contours are processed in spoken stimuli that are related to *ASPM*. All listeners were genotyped and for purposes of analysing the results they were divided into three groups according to the number of copies of the new version of *ASPM* they had in their genotype – zero, one, or two.

In one task, listeners were required to make phonetic judgements about the pitch contours on vowel-only syllables pronounced with rising, level, or falling tone. In this task, the results showed clearly that the more copies of the new version of *ASPM* the listeners have, the better they are at identifying the phonetic properties of the pitch contour in spoken syllables. There was another task as well, but the results were less clear, and I don't have time to discuss this in enough detail. But the phonetic identification task seems to show clearly that ASPM does have an influence on the perception of pitch in spoken stimuli, in line with the Dediu and Ladd speculations. However, Wong and colleagues maintain that the effect is the ***opposite*** of what would be predicted by Dediu and Ladd. Dediu and Ladd showed that the ***old*** version of *ASPM* is associated with tone languages, but Wong et al. show that it's the ***new*** version that has an effect on the processing of pitch contours.

What I want to suggest is that Wong et al's interpretation misses something crucial, and that there's no contradiction here. What their results show is ***not*** that the listener is somehow better at ***tone*** – they're better at ***pitch contours***. That is, the listeners with more copies of the new version of *ASPM* are ***making a clearer separation*** between the pitch contour and the segmental identity of the syllable. That's exactly what you have to do to speak a language with European style intonation – you have to abstract the segments away from the pitch contour – but it's exactly the opposite of what you have to do to speak a language like Chinese, where you want to keep pitch and segments integrated. So I believe that the Wong et al. findings actually do support Dediu and Ladd's speculations. And if this is the right way to think about this, then the MF data seem to be relevant here as well – that is, the MF findings demonstrate that people have different ways of processing acoustic signals to arrive at percepts of both the pitch and the timbre or spectral composition.

However, even if you're convinced by what I have said so far, it raises a key typological question. The biggest problem with my reading of Wong et al. comes from African languages – and some Mexican languages as well. In East and SE Asian tone lgs., what I just said is more or less true – you want to keep pitch and segmentals integrated. But in many African tone lgs., the pitch contours are clearly separate from the segmentals, in ways that are well known. They slide around autosegmentally, and many speakers have no trouble thinking about the pitch contour of an utterance separately from the segmentals, and so on. That is, in a sense ***African*** and ***European*** lgs. are ***more alike*** in the relation between ***pitch and segmentals***, whereas African and ***E Asian*** lgs. are more alike in the relation between ***pitch and lexical/grammatical meaning***. I simply don't know how to fit these differences into a satisfying story about Wong et al.'s results, or into a satisfying typology of linguistic pitch.

So we need several things to make progress:

First, we really need to find out more about Africans' pitch perception. There are some fairly well-documented reports of differences between Caucasians and East Asians with regard to pitch perception, but I'm not aware of any research on Africans. I think everyone is becoming more aware of the extent to which what we think we know about how people think and perceive is actually knowledge about how white undergraduates think and perceive, and that in the long run we need to expand our horizons. Given the clear evidence for individual differences in pitch perception and the possibility that these are genetically influenced to at least some extent, I think African pitch perception is an obvious place to start.

Second, we need a focused genetic study to see if *ASPM* is related to missing fundamental perception. Dan Dediu is already working on this and related studies at the Max Planck Institute in Nijmegen. Things move slowly, but I hope it won't be too long before we have some findings on this as well, and on how it ties in with other language-related tasks.

And third, we need more theoretical and descriptive thinking about the typology of the way lgs use pitch. Whatever you think of the idea of linguistic typology, pitch is one area where there are conspicuous differences that seem to demand some sort of typological account. But there are many many diverse and contradictory ideas out there about how all the non-segmental properties of language relate to each other, and unquestionably we need new insights. I'd be very interested to hear your reactions to my suggestion about the relation between pitch and segments in typical African, East Asia, and Eurasian languages.

References

Dediu, D. and Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. PNAS 104: 10944-10949.

Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson JR, Vaez-Azizi LM, Tishkoff SA, Hudson RR, Lahn BT (2005). Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science 309:1717-1720.

Ladd, D. R., Turnbull, R., Browne, C., Caldwell-Harris, C., Ganushchak, L., Swoboda, K., Woodfield, V., and Dediu, D. (2013). Patterns of individual differences in the perception of missing-fundamental tones. Journal of Experimental Psychology: Human Perception and Performance, vol. 39. DOI: http://dx.doi.org/10.1037/a0031261.

Lahn group: see Mekel-Bobrov et al., Evans et al.

Mekel-Bobrov N, Gilbert SL, Evans PD, Vallender EJ, Anderson JR, Hudson RR, Tishkoff SA, Lahn BT (2005). Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. Science 309:1720-1722.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., … Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. Nature Neuroscience, 8, 1241-1247.

Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R., Seither, S., & Lütkenhöner, B. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. Journal of Experimental Psychology: Human perception and performance, 33, 743-751.

Smoorenburg, G. (1970). Pitch perception of two-frequency stimuli. Journal of the Acoustical Society of America, 48, 924-942.

Wong, P. C. M., Chandrasekaran, Bh., and Zheng, J. (2012). The derived allele of ASPM is associated with lexical tone perception. PLoS-ONE 7(4): e34243.