Differential Analysis of Lexical Pitch in Accent and Tone Languages

Ettien Koffi
St. Cloud State University, enkoffi@stcloudstate.edu

Follow this and additional works at: https://repository.stcloudstate.edu/stcloud_ling
Part of the Applied Linguistics Commons

Recommended Citation
Available at: https://repository.stcloudstate.edu/stcloud_ling/vol7/iss1/8
Differential Analysis of Lexical Pitch in Accent and Tone Languages

ETTIEN KOFFI

ABSTRACT

According to the Critical Band Theory, the auditory perception of F0 data is the same for all human beings. However, when F0 signals are transferred through the auditory cortex to specialized areas of the brain, they are perceived and processed differently, depending on whether the language is tonal or accentual. In tone languages, F0 data appears to be processed in Heschl’s gyrus (Schneider 2005, Bendor 2012), whereas in accent languages, it appears to be processed in the planum temporale (Binder et al. 1996). Furthermore, in accent languages, F0 signals are computed on a nominal scale, but in tone languages, a logarithmic scale is used (Wightman 1973, Speaks 2005). These insights support the long-held linguistic view that accent and tone languages are prosodically different. Terms such as strong/weak or stressed/unstressed are used to describe pitch variations in accent languages, whereas in tone languages, the terms used are extra low, low, mid, high, and extra high. Current research on language and the brain suggests that the differences between these two types of languages may be the result of differences in tonotopic mapping, autocorrelational algorithms, and the scales on which pitch is computed. Due to these differences, it is not advisable to apply the same interpretive framework in analyzing pitch variations in accent and tone languages. Examples will be provided from English and Baule, a West African language, to underscore the pitfalls of doing so.

1.0 Introduction

This paper investigates issues at the intersection of acoustic phonetics and neuroanatomy with regard to the intelligibility of lexical pitch. We know that the articulatory apparatus that humans use to produce pitch is the same. We also know that the auditory processes that they deploy to perceive pitch is the same. Identical also are the neural transduction system and the auditory pathway that underlie the processing and intelligibility of pitch. Given these undeniable similarities, is it justified for linguists to make a distinction between accent and tone languages? I argue in this paper that the distinction has merits and that it is based, first, on the areas of the brain where lexical pitch is processed, and secondly, on the nature of the autocorrelational analyses that the cerebral cortex performs unbeknownst to the hearer. The arguments in support of this view are made in four separate sections of the paper. The first section addresses definitions and a general review of the literature. The second focuses on the similarities in production, in neural transduction, in critical bands, and in auditory pathway processing. The third discusses the neuroanatomical findings in support of the Duplex Model of Pitch Perception (DMPP). The last section highlights the dos and don’ts of pitch analysis in accent and tone languages.

---

1 I wish to thank the participants of the 174th meeting of the Acoustical Society of America meeting in New Orleans, Dec 03-08, 2017 for insightful comments and observations during the poster presentation that was the precursor of this paper.
2.0 Definitions and a Quick Literature Review

The study of pitch has a long history. Some trace it back to Humboldt and others go as far back as to Pythagoras (Brownell 2017:20). Yet contemporary understanding of pitch owes much to Fletcher and Békésy for their separate but complementary works on the physics and physiology of the basilar membrane (Brownell 2017:20, 23, Heller 2013:423, Yost 2015:46-53). Pitch research has continued unabated since. There are now sophisticated tonotopic inquiries that bring together physicists, audiologists, otolaryngologists, and auditory neuroscientists. The important findings in these areas have not yet penetrated the field of linguistics at large. Consequently, the linguistic understanding of how pitch is mapped into the auditory system and later on into the cerebral cortex is still rudimentary. It is fair to say that many linguists (i.e. those with whom I go to linguistics conferences) are more interested in the functional uses of pitch in language than its neurolinguistic bases. However, it will be argued in this paper that even a passing interest in the neuroanatomical discussions of pitch can help in understanding the prosodic distinction between accent and tone languages and ground the distinction on a solid scientific footing.

Traditionally, the distinction between accent and tone languages has rested mostly on the functional uses of pitch at the lexical and grammatical levels. However, the functional criterion is not without its shortcomings. Why is English classified as an accent language, and not a tone language like Anyi, even though the former also has lexical minimal pairs just like the latter? Examples of pitch-related lexical minimal pairs in English are <offEnse> vs. <Offense>, <defEnse> vs. <dEfense>. English also has pitch-based grammatical distinction between verbs and nouns such as <addrEss> vs. <Address>, and <resEArch> vs. <rEsearch>, just like Anyi has pitch-related grammatical distinctions in conjugation. In reality, according to Bolinger (1978:479), a clear-cut distinction between accent and tone languages is hard to make:

So, it is hard to make a clear distinction between tone and accent. If accent is primarily a matter of pitch, it might seem to differ only qualitatively from tone – one tonal distinction per word, for example, instead of several; or grammatically – the word rather than a syllable (or a vowel) carrying the pitch change. […] The difference is that high pitch for accent affects its environment, whereas high tone is affected by its environment. For example, high tone may be (in fact tends to be) shorter than low tone, whereas high accent tends to be longer; accented syllables tend to have full vowels and unaccented ones reduced vowels, but tone does not affect vowel quality; accent affects quality of a consonant, but quality of a consonant may well affect tone; etc. It is logical that accent and tone should differ in this way, simply because of the function of accent, which is generally to highlight an element in focus. It has to be able to override opposition, and accordingly calls upon other acoustic means besides pitch, especially duration. Lexical tone is only one of the phonemes making up a word, and context can easily make it redundant. Accent is directly meaningful, tone is only distinctive.

Even the claim that accent languages have schwas on the weak syllable is questionable. There are numerous disyllabic words such as <maybe>, <between>, <because>, etc., in which there is no schwa in any of the vowels. None of the putative distinctive features mentioned in the quote is sufficiently discriminating. No criterion used to differentiate between accent and tone languages can withstand scrutiny. It is nearly impossible to differentiate satisfactorily between accent and tone languages because the very notion of “pitch” on which linguists rely to do the classification
is slippery and intractable. The American National Standards Institute (ANSI) tried to define “pitch” in 1973 as “that attribute of the auditory sensation in terms of which sounds may be ordered on a scale extending from low to high”. However, this definition has been criticized as inadequate for omitting two pitch levels: extra low and extra high (Houtsma 1995:267). The International Phonetic Association (1999:14) recognizes five pitch registers in human languages – extra low, low, mid, high, extra high – but the ANSI’s definition recognizes only two or three pitch levels.

These difficulties notwithstanding, linguists have known for a very long time that accent and tone languages are prosodically dissimilar. Different terms are used to refer to the linguistic manifestations of pitch in these languages. In accent languages, labels such as stressed/strong are used to describe the syllable that has a higher pitch. The syllables that do not have a high pitch are referred to as unstressed/weak. Such terms are never used in the description of pitch variations in tone languages. The labels commonly used to describe pitch variations in tone languages are extra low, low, mid, high, and extra high. Do the terminological differences underscore a difference in the neural processing of pitch in accent and tone languages? I believe they do, and I will demonstrate it by appealing to the Duplex Model of Pitch Perception (DMPP) proposed by Licklider as far back as 1951. He argued that in perceiving pitch in English (and presumably other accent languages), the pitch that is perceived by hearers is different from the pitch of the acoustic signals that is processed by the cochlea. The differences between the two lie in the fact that once pitch data enters the brain, autocorrelational analyses are performed on neural transmissions. Furthermore, accumulated findings from nearly 20 years of auditory neuroanatomy research and tonotopic mapping suggest that pitch in accent and tone languages are very likely processed in different parts of the cerebral cortex. However, before tackling the differences, we will first highlight three areas where pitch is produced and processed identically in accent and tone languages.

2.1 Anatomical Similarities in Pitch Production

The phonetic theory that best explains how humans produce speech similarly is the Source-filter theory. Kent and Read (2002:305) define it simply as “a theory of acoustic production of speech that states that the energy from a sound source is modified by a filter or sets of filters. For vowels, the vibrating vocal folds usually are the source of the sound energy and the vocal tract resonances (formants) are the filters.” One of the main tenets of this theory is that the vocal tracts that human beings use to produce pitch are anatomically similar in configuration and in length. Information collected from Stevens (2000:9, 13, 24, 25) is displayed in Table 1 to underscore the anatomical similarities of the organs that affect the production of pitch and other speech sounds:

<table>
<thead>
<tr>
<th>Articulatory Characteristics</th>
<th>Adult female</th>
<th>Adult male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocal tract length</td>
<td>14.1 cm</td>
<td>16.9 cm</td>
</tr>
<tr>
<td>Pharynx length</td>
<td>6.3 cm</td>
<td>8.9 cm</td>
</tr>
<tr>
<td>Oral cavity length</td>
<td>7.8 cm</td>
<td>8.1 cm</td>
</tr>
<tr>
<td>Vocal fold length</td>
<td>1.0 cm</td>
<td>1.5 cm</td>
</tr>
<tr>
<td>Length of the trachea</td>
<td>2 to 4 cm</td>
<td>2 to 4 cm</td>
</tr>
<tr>
<td>Vocal tract volume (closed mouth)</td>
<td>130 cm³</td>
<td>170 cm³</td>
</tr>
<tr>
<td>Vocal tract volume (mouth open 1 cm wide)</td>
<td>150 cm³</td>
<td>190 cm³</td>
</tr>
</tbody>
</table>

2Thai has five pitch contrasts for the word /na/: 1) low: [nà:] (a nickname), 2) mid: [nā:] (rice paddy), 3) low: [ná:] (young maternal uncle), 4) falling: [nâ:] (face), and 5) rising: [nǎ:] (thick) (Fromkin et al. 2014:212).
Table 1: Estimated Vocal Track Length

<table>
<thead>
<tr>
<th>Volume of the tongue</th>
<th>90 cm³</th>
<th>110 cm³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lip opening (horizontal: from corner to corner)</td>
<td>10 to 45 mm</td>
<td>10 to 45 mm</td>
</tr>
<tr>
<td>Lip opening (vertical: from upper to lower lip)</td>
<td>5 to 20 mm</td>
<td>5 to 20 mm</td>
</tr>
</tbody>
</table>

The F0 ranges that human speech organs produce are the same for the same genders and age groups, irrespective of whether they speak an accent or a tone language. Years of cross-linguistic studies of F0 in hundreds of languages have yielded the same results:

The frequency of vibration of the vocal folds during normal speech production is usually in the range of 170-340 Hz for adult females, 80-160 Hz for adult males, and 250-500 Hz for younger children. The frequencies can extend well beyond these ranges for the singing voice (Stevens 1998:1232).

This means that, as far as F0 is concerned, males who speak accent languages sound exactly like males who speak tone languages, and vice versa. The same goes for females and children. In other words, pitch ranges do not vary as to whether or not a person speaks an accent language or a tone language. Fry (1979:68) lists the F0 averages for men, women, and children respectively as 120 Hz, 225 Hz, and 265 Hz.

2.2 Similarities in Auditory Frequency and Transduction

The anatomy and physiology of the ear is the same for all humans. Structurally, the ear has three main parts: the external ear, the middle ear, and the inner ear. The external ear includes the pinna and the auditory canal. The former gathers sound waves and funnels them through the latter, which is 26 mm long and 7 mm wide. Ballanchanda (1997:412, 416) notes that because the ear canal is not a uniform tube from beginning to end, frequencies are amplified in different sections along the auditory canal. The tympanic membrane marks off the end of the auditory canal and the beginning of the middle ear. Wilson (2015:14-15) observes that the tympanic membrane (i.e., eardrum) also amplifies incoming sounds. The middle ear occupies a very small space: 6 mm at the bottom, 4 mm in the middle, and 6 mm at the top. It contains the ossicles, three interconnected bony structures that are the smallest bones in the human body: the hammer, the anvil, and stapes (also known as stirrup). The hammer is directly connected to the eardrum. Sound waves from the ear canal impinge on the tympanic membrane which causes the hammer to hit the anvil, which in turn, passes the vibrations on to the stapes. The inner ear contains the cochlea, a snail-like structure that has been nicknamed the frequency analyzer of the ear. Inside the cochlea lies the basilar membrane. Fletcher’s groundbreaking experiments in the 1940s and Békésy’s clinical demonstrations in the 1950s have advanced and enhanced contemporary understanding of the role that the basilar membrane plays in frequency analysis and perception. It is 30 to 35 mm long in adults and is compartmentalized into critical bands. Each critical band encompasses an area where frequencies of the same range are perceived. Humans generally can perceive frequencies that range from 20 to 20,000 Hz. However, the frequencies involved in the production of speech sounds (vowels and consonants) are ≤ 5,000 Hz (Heller 2013:474, 479). For F0/pitch, the frequencies range from 75 to 500 Hz for all human beings, regardless of whether their native language is accentual or tonal. This explains why Praat has this range as its default settings for pitch analysis.
The process whereby the acoustic signals are converted into neural pulses is called transduction. Some 3,000 inner and 12,000 outer hair cells line up along the basilar membrane and convert acoustic signals coming from the outer and middle ears into electrical signals that they pass on to the auditory nerve (See Brownell (2017:20-27) and Lewis (2016:40-48) for a detailed description of the action of hair cells.) Figure 1 explains pictorially how this happens. The portion from the Base to the Apex describes the frequencies perceived by the basilar membrane with specified critical bands. These critical bands and their corresponding frequencies are the same for all human beings regardless of whether they speak an accentual or a tonal language.

![Figure 1: Audibility Range in the Frequency Domain](image)


The transduction process is indicated by the arrows pointing towards the auditory nerve fibers. Stevens (2000:211) reports that there are 30,000 of them. They travel along the auditory pathway and carry acoustic information to the brain (see description below). The transduction system is also similar for all human beings, irrespective of whether their native language is accentual or tonal.

### 2.3 Similarities in Auditory Pathway and Pitch Processing

The transduced acoustical signals are processed as nerve impulses and travel alongside the auditory pathway up to the thalamus, nicknamed the “gatekeeper” of various sensory inputs that the brain processes. Amerman (2016: 434) describes its role in sensory perception as follows:

> The thalamus consists of two large, egg-shaped masses of gray matter that together make up 80% of the diencephalon. […] The thalamus nuclei receive input from multiple sources, including the cerebral cortex, the cerebellum, the basal nuclei, structures of the limbic system, and the sensory system (except the sense of smell), and their main output travels the cerebral cortex. The thalamus is literally the ‘main entrance’ into the cerebral cortex—nearly all information destined for the cerebral cortex must first pass through the thalamus.
This allows the thalamus to control which information reaches the cerebral cortex and where the information is sent, which means that ultimately the thalamus regulates cortical activity.

Once the thalamus receives F0 data from basilar membrane via the auditory pathway, it dispatches it to the “pitch centers” of the brain in the primary auditory cortex. Amerman (2016:573) sketches the process as follows:

Step 1: Auditory signals travel through axons of the cochlear portion of the vestibulocochlear nerve to the cochlear nuclei at the medulla-pons junction.
Step 2: Axons from the cochlear nuclei contact the superior olivary nucleus in the pons.
Step 3: Auditory stimuli are then sent to the inferior colliculus of the midbrain.
Step 4: The auditory stimuli are relayed to the medial geniculate nucleus of the thalamus.
Step 5: The thalamus sends signals to the primary auditory cortex in the superior portion of the temporal lobe.

Figure 2 helps to visualize the process better, because, as the saying goes, a picture is worth a thousand words:

![Figure 2: The Auditory Pathway](image)


The auditory pathway is the same for all human beings, regardless of whether their native languages are accentual or tonal. Amerman (2016:573) also notes that “The primary auditory cortex has connections with other parts of the temporal lobe that are specialized for language, and
with the limbic system for emotions and memory.” So far, we have highlighted the areas of similarities between accent and tone languages with regard to pitch processing. In the upcoming sections, we turn our attention to the neuroanatomical factors that cause speakers of some languages to perceive pitch as accent, while others perceive it as tone.

3.0 The Duplex Model of Pitch Perception

The differential perception of pitch as accent or as tone has a lot to do with the neural and cortical system of the brain and also with algorithms and scales that the brain uses to compute incoming frequency data. As far back as 1951, Licklider proposed the Duplex Model of Pitch Perception (DMPP) to explain the “pitch problem” that scholars had been faced with for a long time. The conundrum for perceiving frequency and translating it into lexical pitch can be stated as follows: if all aurally healthy individuals have structurally similar basilar membrane and the same critical band responses, how can people perceive very minute changes in frequency (i.e. 0.3%) as to whether a syllable is stressed or not? Faced with this dilemma, we could either jettison the critical band theory altogether, or we could find an alternative explanation. The anatomical and physiological bases of frequency analysis in the cochlea are too strong and too important to ignore. Therefore, Licklider (1951:128) proposed a theory according to which pitch is perceived in two stages. Here is his explanation:

In the theories of pitch perception now widely supported, pitch is regarded as a unitary attribute of auditory experience. There is good evidence, however, that there are actually two pitch-like attributes, and it is reasonable to suppose that the duplexity of pitch is a reflection of duplexity in the auditory process. The first step in the process is analysis in frequency, performed by the cochlea, which distributes stimulus components of various frequencies to spatially separated channels. The second step, according to the scheme postulated here, is autocorrelational analysis, performed by the neural part of the auditory system, of the signal in each frequency channel.

The DMPP is widely accepted now as a viable solution to the pitch problem. It is behind many explanations whether it is referred to explicitly or not. Bendor’s (2012:743) analysis assumes DMPP even though he does not mention it explicitly anywhere in his paper:

Because spectrally different sounds can have the same pitch, the frequency-based neural representation of a sound in the auditory periphery does not explicitly encode pitch, and pitch must be computed at some later stage in the auditory pathway.

Bendor is essentially saying here what Licklider said nearly 70 years ago. Tonotopic evidence based on PET (Positron Emission Tomography), MEG (magnetoencephalography), MRI (Magnetic Resonance Imaging), fMRI (Functional Magnetic Resonance Imaging), EPI (Echo-Planar Imaging), and other sophisticated and state-of-the-art tools such as those described in Lee (2017:36-43) have confirmed that pitch is perceived in two main areas of the primary auditory cortex. However, before delving into some details, we must acquaint ourselves broadly with the main cortices of the brain by taking a look at Figure 3:
Figure 3 gives us a broad view of the area of the auditory cortex, the area of the brain where speech sound (all sounds for that matter, but we are interested here only in speech) is processed. The auditory cortex has two sub-areas: the primary auditory cortex and the auditory association cortex. We are primarily interested in the latter, specifically the pitch perception centers situated in Heschl’s gyrus (HG) and in the Planum temporale (TP), shown in Figure 4:

The remainder of this paper focuses on the role that HG and PT play in the processing of pitch data found in speech. It will be shown that in tone languages pitch is processed solely in HG, whereas in accent languages, pitch is first processed in HG and then processed again in PT. Mathematical calculations known as autocorrelations intervene in processing pitch information in accent languages, but not in tone languages. The net result is that pitch in accent languages loses spectral resolution, which reduces pitch registers from five to only two.
3.1 Pitch Processing in Heschl’s Gyrus (HG)

The emerging consensus among researchers today is that two “pitch centers” exist in the brain. Both are located in the primary auditory cortex. The first is Heschl’s gyrus (HG), discussed in this section, and the second is the planum temporale (PT), to be addressed in 3.2. The auditory cortex is very small, occupying only 8% of the cerebral cortex (Woods and Alain 2009:407). Recent tonotopic investigations have revealed that the neurons in HG are very sensitive to pitch and respond to frequencies differently depending on which of its three areas is activated. The posterior (towards the back) perceives high frequency pitches, the anterior (towards the front) low frequency pitches, and the middle area mid-frequency pitches (Woods and Alain 2009:408). A very important characteristic of HG that has a direct bearing on pitch perception is that the critical bands from the basilar membrane are replicated and projected directly there, as shown in Figure 5:

![Figure 5: Frequency Mapping of Critical Bands into HG](image)

Bizley and Walker (2010) cite numerous studies showing that HG is indeed an important pitch center. Additionally, Schneider et al. (2005:1245) have found that HG in the right hemisphere is also active in pitch perception:

Thus, the existence of two pitch centers may facilitate the extraction of fundamental pitch in left auditory cortex and spectral pitch in right auditory cortex. Indeed, most professional musicians perceive simultaneously both fundamental and spectral pitch from an ambiguous tone, and the subjective differences are rather relative than absolute. Here, these relative perceptual differences were found to correlate strongly to neural asymmetries, as anticipated by earlier studies on cerebral dominance. Thus, a greater volume on the left may predispose one to hear the F0 in an ambiguous tone, and vice versa, a greater volume on the right may lead to a dominant perception of spectral pitch or single harmonics.

---

3 A similar picture appears in Purves et al. (2012:298, Figure 13.15A). *Neuroscience, Fifth Edition*. Oxford University Press granted me permission to use this picture in oral presentation at the 174th meeting of the Acoustical Society of America in New Orleans, but not for this publication.
However, given that language functions are lateralized mostly in the left hemisphere for 97% of people, most pitch researchers have not investigated vigorously the connection between HG in the right hemisphere and pitch perception. It is not unreasonable to think that some aspects of pitch can be perceived in HG in the right hemisphere because, as has been noted in the literature, emotions are processed there (Amermann 2016:469, Marieb and Hoehn 2013:437, 465).

3.2 Pitch Processing in the Planum Temporale (PT)

The other pitch center that is often discussed is the planum temporale (PT). Binder et al. (1996:1240) conducted a functional MRI imaging study of the brains of 12 participants (six men and six women) and found that the neurons in PT were very sensitive to pitch. They list a number of characteristics associated with PT that qualify it as a pitch center. They note, for instance, that impairments to this area lead to deficits in auditory discrimination and speech comprehension. This is not really surprising since PT, like HG, is also located in the auditory cortex (Hackett 2008:774). Zheng (2009:3079), Saenz and Langers (2014:44-45) note that the exact boundaries of PT are hard to establish because it extends into other association areas. Some portions of PT lie in close proximity to HG, as shown in Figure 4 above. Yet, PT is anatomically different from HG. It is easily identifiable by its triangular shape (Bendor 2012:745). PT also has two parts: the anterior PT and the posterior plane (PP). Binder et al. (1996:1245) describe the pitch sensitivity of PT as follows:

The posterior portion of the left PT was more strongly activated by nonlinguistic stimuli (tone sequences) than by words. This preference for nonlinguistic stimuli was even more pronounced in the PP and neighbouring parietal operculum and when subjects performed active tasks involving tone sequence analysis in comparison to analysis of words. [...] yet association of this area and PT with processing of tone sequences was unexpected. These stimuli are in a sense musical, and it may be that the PT, PP (posterior plane) and parietal operculum play some special role in music perception. More explicitly, this role could relate to the psychological phenomenon of pitch perception, which is a visualization of spectral frequency on a spatial continuum from low to high. The perceptually salient positional shifts of the tone stimuli in the pitch domain might therefore elicit processing by neural networks also involved in spatial representation, located in the inferior parietal cortex.

Zheng (2009:3079) notes that PT “represent[s] higher order processing stages in the hierarchically organized auditory processing system.” We interpret “higher order” processing as “additional processing.” In other words, it is quite likely that pitch data from HG is sent on to PT for further processing. Lee (2017:40) seems to be making this very point with regard to active versus selective listening of speech:

Localization analysis of earlier sources suggests that the neural source of these earlier components originates from Heschl’s gyrus … whereas the neural sources of the later components originate from the planum temporale. A possible interpretation is that the

---

4 We do not concern ourselves with the myriads of ways in which human beings, no matter which society they live in, learn early on through socialization that they can manipulate pitch to convey an assortment of human emotions, including, but not limited to, amazement, joy, annoyance, anger, empathy, boredom, affection, and fear.
entire auditory scene is processed by the primary auditory cortex and it is only weakly sensitive to selective attention, whereas the higher order auditory areas in the planum temporale receive the processed neural signals, with the speech streams already segregated.

This supports the argument that pitch data in PT is simplified compared to pitch data in HG. The simplification comes as a result of additional processing. We now turn to the neural algorithms that modulate this simplification.

3.3 Autocorrelational Analyses and Pitch Processing in Accent Languages

Pinpointing the “pitch centers” in the human brain is important. Equally important and probably just as consequential for the distinction between accent and tone languages is understanding how the acoustical signals are exploited as neural impulses and processed later as lexical pitch. Licklider (1951:128) was among the first to hypothesize that the brain performs a spectrotemporal analysis in perceiving pitch. The analysis consists of some autocorrelational algorithms which he described as follows:

The essence of the duplex theory of pitch perception is that the auditory system employs both frequency analysis and autocorrelational analysis. The frequency analysis is performed in the cochlea, the autocorrelational analysis by the neural part of the system. The latter is therefore an analysis not of the acoustic stimulus but of the trains of the nerve impulses into which the action of the cochlea transforms the stimulus. This point is important because the highly nonlinear process of neural excitation intervenes between the two analyses [some portions are highlighted here for emphasis].

DMPP posits that pitch is processed in two stages: first in the cochlea and secondly inside the brain. Licklider referred to the analysis in the cochlea as “frequency analysis” and the analysis inside the brain as “autocorrelational analysis.” Heller (2013:72-3) describes autocorrelational analyses as “a result of high-level data processing in the brain, rather than built-in properties of the receptor in the ear. The sensation of pitch is the end result of our human autocorrelation algorithm.” At first, Licklider’s idea of autocorrelation analyses for pitch extraction received a lukewarm reception. However, things have changed. Rabiner (1977:24) writes, “Although a large number of different methods have been proposed for detecting pitch, the autocorrelation pitch detector is still one of the most robust and reliable pitch detectors.” Kent and Read (2002:97) acknowledge that there are still some problems with autocorrelation, but state that “Despite these difficulties, autocorrelation is one of the more reliable methods for determining fundamental frequency.” Heller (2013:450) issues the following warning to pitch researchers: “Beware of any theory of pitch perception that entirely leaves out autocorrelation.”

As noted in 3.2, many researchers regard pitch perception in PT as the result of a “higher order” processing operation. Better yet, it can be construed simply as additional processing of pitch data from HG which results in a loss of some spectrotemporal resolutions. Autocorrelational analyses are responsible for simplifying the number of pitch registers from five to two. Rabiner and Schafer (1978:151-4) discuss autocorrelation and simplification at length. Readers who want

---

5 Meddis and O’Mard (1997) have proposed “A Unitary Model of Pitch Perception” in which they argue against the Duplex Model. This model can account for pitch perception in tone languages. However, it may fail to account for pitch-accent languages.
to learn more should refer to them. In other words, when the speakers of accent languages hear lexical pitch, unbeknownst to them, they perform autocorrelation analyses that cause them to hear the pitch of syllables as high or low, strong or weak, stressed or unstressed. This is not so for the speakers of tone languages. They process pitch data directly from HG. In so doing, they bypass autocorrelational analyses and process pitch sensitive neurons directly in Wernicke’s area. As a result, these speakers can perceive all pitch levels more accurately than the speakers of accent languages. This claim is supported by experiments in absolute pitch perception studies. It has been reported (Heller 2013:475 and others) for instance, that native speakers of Mandarin are better at perceiving absolute pitch than speakers of American English):

This finding resonates with the claim that the prevalence of AP (absolute pitch) is extremely rare in the U.S. and Europe, with an estimate of less than one in 10,000 in the general population…. For tone language speakers, the neural circuitry for acquiring the association between pitches and verbal labels is already in place, so that they are able to build the ability of acquiring absolute pitch on it with greater success (Peng et al. 2013:458).

For languages with five contrasting pitch registers: extra low, low, mid, high, and extra high, native speakers can perceive all the variations clearly. Speakers of languages that have two or three pitch levels can readily perceive the contrast between low, mid, and high registers well.

3.4 Temporal and Distal Characteristics

Another factor that plays a role in the simplification of F0 frequencies in accent languages is the additional time pitch neurons take to transfer the data to PT. This is referred to as delay (distance) in the literature. Delay speaks to the distance that the electrochemical impulses have to travel (Bidondo et al 2013). There are two important truisms in signal processing: (1) the closer a signal is to the emitting source, the stronger and the richer it is perceived; (2) the less time it takes to transmit a signal to its final destination, the stronger and the richer it is. Both have implications for the perception of pitch as accent or tone in world languages. Licklider (1951:129) touches on these two aspects in a footnote saying, “For the sake of simplicity, we assume that delay in time is proportional to the distance traversed in the neural tissue.” Later, on page 132, he notes that “the autocorrelator should be as near the cochlea\(^6\) as possible so that it may operate upon the signal before temporal resolution is lost.” Speakers of tone languages process pitch registers in all their richness because transmission is sent directly to Wernicke’s area without further processing or delay. In other words, they have the richness of the pitch data to operate with.

The contention that speakers of tone languages do not lose a lot of frequency resolution in processing F0 data may be supported by Myers et al. (2009) and Myers (2017:37). She explains that “As the neural processing advances away from the superior temporal gyrus to other areas of the temporal lobe and towards the left frontal brain areas, the representation of sounds appears to lose some of the fine-grained acoustic details.” She provides the picture below to illustrate this general principle:

\(^6\) Without having the benefits of modern tonotopic mapping techniques, Licklider could not have known that the frequencies and critical bands found in the cochlea are projected in HG. This does not invalidate his model in any way.
The caption underneath the original image reads as follows:

Sensitivity to speech sounds shows a gradient of processing. Areas in Heschl’s gyrus (not pictured) and the part of the superior temporal gyrus highlighted in green display sensitivity to the fine-grained acoustics of many different speech sounds. As processing spreads away from those central locations—to the yellow-highlighted areas of the superior temporal gyrus and frontal region of the brain—the neural signal tends to embody only the acoustic differences that listeners use to distinguish between words (Myers (2017:35)

To be sure, Myers’ article deals only with the neural processing of segments, not of suprasegmentals. Yet the same principle of spectrotemporal resolution may apply to pitch perception. Speakers of accent languages lose sensitivity to fine-grained pitch register data because they process pitch twice: once in HG and the second time in PT.

3.5 Autoscaling Analyses in Pitch Perception

Another important attribute of pitch perception is the scale on which it is weighed. Wightman’s (1973) mathematical model of pitch perception discusses several weighted algebraic functions. However, since the mathematical jargon is very advanced, we fall back on Speaks’ (2005:97-100) simpler discussions of the scales involved in processing the frequencies involved in speech. He deals with four kinds of scales, but two, the nominal and the logarithmic scales, are relevant to the analysis at hand. “A nominal scale is one on which things can be sorted into different categories by observing that one object is the same as or different from another. That is the only requirement.” A logarithmic scale, on the other hand, operates differently. It is a ratio scale that consists of “successive units that are generated by multiplying (or dividing) each successive number along the scale by the base.” The two scales are used by hearers in processing lexical pitch data. As noted earlier, native speakers of accent languages weigh lexical pitch on a binary fashion. They perceive syllables either as strong or weak, stressed or unstressed. In so doing, they rely on a nominal scale to process pitch. Speakers of tone languages, on the other hand, can discriminate
between up to five different levels of lexical pitch. In so doing, they weigh frequencies on a logarithmic scale. This is consistent with Saenz and Langers’ (2014:47) finding that a logarithmic scale is used to weigh and process pitch information in HG.

The following metaphor may be used to explain and illustrate how accent and tone languages process pitch. For the sake of this analogy, let’s assume that speakers of accent languages process pitch on a Pass/Fail scale. This scale is binary: pitch is either high or low. Therefore, in any given word, syllables are either strong or weak, stressed or unstressed. Speakers of tone languages, on the other hand, process pitch on a logarithmic scale that goes from F (extra low), D (low), C(mid), B (high) to A (extra high) by analogy with the academic grading scheme. The pitch perception scales and the academic grading scales are similar in many respects. Pitch perception in accent languages is similar to the Pass/Fail grading system. Pitch perception in tone languages is similar to the shaded grading system that discriminates between five letter grades. This system provides teachers and other interested parties with more details about a student’s performance because it has gradient values. In like fashion, speakers of tone languages are attuned to the richness of F0 modulations because they depend on the logarithmic scale for lexical and/or grammatical meaning. The use of two different scales to perceive pitch in accent and tone languages is illustrated in 4.1 and for 4.2 when we compare English and Baule.

4.0 Pitch Analysis in Accent and Tone Languages

Impressionistic assessments of lexical and grammatical pitch have reigned supreme in linguistic analyses. However, in recent years, especially since the advent of Praat, thanks to Boersma and Weenink (1995), many linguists are basing their statements of pitch registers on acoustic measurements. This has led to situations where measurements do not correspond to the pitch level at which they are pegged. Moreover, even though linguists know that accent and tone languages are functionally different, many still interpret pitch measurements in tone languages in light of pitch measurements in accent languages or vice versa, as in the case of Gandour (1978). This should not be, because as noted above, in processing pitch in accent languages, additional autocorrelational functions are performed, but not so in tone languages. Moreover, two different autoscaling methods are used. The following sections highlight how lexical pitch should be interpreted in these two prosodically distinct types of languages.

4.1 Pitch Analysis in Accent Languages

One of the most important studies on the perception of lexical pitch in accent languages is the one conducted by Fry in 1958 when he set out to determine which of the three acoustic correlates of stress was the most robust in English:

The effect of pitch on perception of stress is generally held to be that a higher pitch produces an impression of greater stress. This experiment was designed to test first the hypothesis that if two syllables differ in fundamental frequency, the syllable having the higher frequency is more likely to be judged as stressed (Fry 1958: 151).

Fry’s overall finding is summed up as follows:

---

7 For a short of history of Praat, see https://en.wikipedia.org/wiki/Praat.
Change in fundamental frequency differs from change in duration and intensity in that it tends to produce an all-or-none effect, that is to say *the magnitude of the change seems to be relatively unimportant while the fact that a frequency change has taken place is all important* [Italics added for emphasis] (Fry 1958: 151).

The Just Noticeable Difference (JND) in the frequency domain for lexical pitch perception is taken to be 0.3%. Young (2011:609) explains it as follows:

The perception of frequency is called pitch. Most of us have excellent relative pitch, which means that we can tell whether one sound has a different pitch from another. Typically, we can discriminate between two sounds if their frequencies differ by 0.3% or more. For example, 500 and 501.5 Hz are noticeably different. Pitch perception is directly related to frequency and is not directly affected by other physical quantities such as intensity.

The JND of 0.3% threshold applies mostly in laboratory settings where clicks and pure tones are used to establish the baseline thresholds for pitch perception (Purves et al. 2012:299). However, for the perception of lexical pitch in the “real” world, Lehiste (1970:64) and a number of phoneticians cited in Gandour (1978:57) use the JND of 1 Hz. In other words, in disyllabic or multisyllabic words, if the nucleus of one syllable is 1 Hz higher than the nuclei of all other syllables, the one with the highest pitch is deemed to be stressed or strong, while another nucleus or nuclei is/are unstressed or weak. In accent languages, once the JND threshold of 1Hz is reached, the auditory-perceptual system stops computing pitch differences, and the syllable with the highest F0 is deemed the stressed/strong one. All others are weak or unstressed. Pitch is therefore perceived categorically in accent languages in a binary fashion. This is what Fry meant by “an all-or-none effect.” In metrical phonology also, stress is classified in a binary fashion: a syllable is either stressed or unstressed (Kenstowicz 1994:553, 596, 599; Goldsmith 1990:172, 183-5). This is the by-product of autocorrelational and autoscaling analyses. As a result, speakers of accent languages process pitch on a nominal scale in a binary fashion.

Nearly every linguistic book defines lexical stress the same way Fromkin et al. (2014:254) do. They state that “To produce a stressed syllable, one may change the pitch (usually by raising it), make the syllable louder, or make it longer. We often use all three of these phonetic means to stress a syllable.” This definition can be translated acoustically into JND thresholds as follows. A syllable is deemed stressed if and only if:

- The F0 of its nucleus is \( \geq 1 \) Hz higher,
- the intensity of its nucleus is \( \geq 3 \) dB louder, or
- the duration of its nucleus is \( \geq 10 \) ms longer than any other nucleus/nuclei in the same word.

---

8 Ladefoged (1996:77-78) reports that laboratory tests have shown that the average JND of pitch perception is 2 or 3 Hz for frequencies below 1000 Hz. However, since the pitch range in male speech is between 80 to 160 Hz, and that of female speech is 170 to 340 Hz, the JND in male speech is 0.24 to 0.48 Hz, and that of female speech is 0.51 to 1.02 Hz. Phoneticians are therefore justified in using the JND of 1 Hz for both male and female speech to make calculations easier. Heller (2013:473) reports that for 200 Hz, the JND is 3 Hz. However, this seems a little high. It is unclear from the context whether or not Ladefoged and Heller were referring to the perception of pitch in human speech.

9 No effort is made here to justify or explain these JNDs. They have been used in acoustic phonetic research and audio engineering applications for nearly 50 years. Readers who wish to know more are invited to refer to Koffi (2017) where these and other thresholds are discussed and applied.
A twenty-year old male speaker from Central Minnesota was recorded reading the six words in Table 1 taken from Fromkin et al. (2014:270). In these examples, the stress falls on the penultimate syllable. Since intensity and duration are not the focus of this paper, we will ignore them in the discussions even though their measurements are displayed in the table.10

<table>
<thead>
<tr>
<th>N0</th>
<th>Correlates</th>
<th>F0</th>
<th>Intensity</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Word</td>
<td>σ 1</td>
<td>σ 2</td>
<td>σ3</td>
</tr>
<tr>
<td>1</td>
<td>astónish</td>
<td>113Hz</td>
<td>115HZ</td>
<td>88Hz</td>
</tr>
<tr>
<td>2</td>
<td>éxit</td>
<td>114Hz</td>
<td>95Hz</td>
<td>89Hz</td>
</tr>
<tr>
<td>3</td>
<td>imágine</td>
<td>113Hz</td>
<td>115Hz</td>
<td>89Hz</td>
</tr>
<tr>
<td>4</td>
<td>cáncel</td>
<td>119Hz</td>
<td>91Hz</td>
<td>92Hz</td>
</tr>
<tr>
<td>5</td>
<td>elícit</td>
<td>120Hz</td>
<td>90Hz</td>
<td>63dB</td>
</tr>
<tr>
<td>6</td>
<td>prácitce</td>
<td>120Hz</td>
<td>90Hz</td>
<td>63dB</td>
</tr>
<tr>
<td>7</td>
<td>Rankings</td>
<td>5/6</td>
<td>4/6</td>
<td>5/6</td>
</tr>
</tbody>
</table>

Table 2: Pitch Measurements in Accent Languages

In the word <imágine>, the nucleus of the syllable <má> is deemed stressed, while the vowels <i> at the beginning of the word and in <gine> are not. The nucleus of the syllable <má> is stressed because its F0 is at least 1 Hz higher than the nuclei in the two other syllables. The same is true for the nuclei in the word <astónish>. This exemplifies how speakers of accent languages such as English encode lexical stress. This is also the way they perceive pitch on a nominal scale, as was mentioned previously.

4.2 Pitch Analysis in Tone Languages

The tonotopic findings in 3.1 indicate that HG is a pitch center. Since the critical bands of the basilar membrane are projected exactly as they are into HG, we deduce that HG is the pitch center from which speakers of tone languages process pitch information (Saenz and Langers 2014:47). Consequently, the Critical Band Theory (CBT) is the best method for correlating F0 measurements in tone languages with their corresponding pitch registers. The advantages of using CBT were amply discussed in Koffi (2017) and will not be repeated here, except to display the correspondence between F0 measurements and their corresponding pitch registers in Tables 3 and 4.11

---

10 We note in passing that this speaker uses F0 and duration equally (83.33%) to encode lexical stress in disyllabic words. He ranks his correlates as follows: F0 (83.33%) = Duration (83.33%) > Intensity (66.66%).

11 Male F0s are the default. One derives female F0s by adjusting male F0s upward by 50% (see Koffi 2017:157 for details). Technically, female F0s are between 1.2 to 1.7 times higher than male frequencies (Kent and Read 2002:191). However, to make calculations easier, it is customary to multiply male F0 by 1.5 times. It is the same as adjusting male F0s upward by 50%. Frequencies can be calculated in another way. Heller (2013:474) notes that the threshold for pitch perception in any given critical band is 1/30 for frequencies between 20 and 10,000 Hz. I prefer the information in Tables 2 and 3 because they make correlating F0 data and pitch registers easier and more straightforward. It is also worth noting that the speakers of many African languages can discriminate only between low, mid, and high pitch registers because many of their languages have only a phonemic contrast between low and high tones.
It is important to note that some male speakers produce F0 ranges that are higher than the typical male range (80 Hz to 160 Hz), as mentioned on page 113. There are several reasons why this may happen. Such men may have a slightly shorter vocal track (see Table 1 for measurements), they may be men of small frame, they may be heavy smokers, they may use specific phonation types, or they may have some speech idiosyncrasies. In such cases, it is acceptable to use women’s critical bands to account for male pitch registers. There are also women whose F0 range (170 Hz to 340 Hz) is lower than that which is typically associated with female speech. In such cases, it is acceptable to use men’s critical bands to describe female pitch registers. Since we are not interested in the speech acoustics of pre-pubescent speakers, we do not concern ourselves with their critical bands in this paper.

5.0 The Dos and Don’ts of Pitch Analysis

The JND threshold of 1 Hz used commonly for pitch perception yields paradoxical results when applied to tone languages. This raises questions about Gandour’s (1978) findings regarding F0 measurements and pitch register levels in Thai. A case in point is Sentence 1a from Ahoua (1996:138). Pages 138-144 of his book contain 27 tables of Baule sentences parsed according to their grammatical mood, pitch registers, and F0 measurements. The dialect of Baule under consideration is 84% mutually intelligible with Anyi Morofou, my native language (Burmeister 1981:4). Sentence 1a was produced by Speaker 1 (K.A.):

Sentence 1a: [n seli ke bla]
Indicative Mood: I said that come (I told him/her to come)
F0 Measurements: 105 100/92 70 100/135
Pitch Register: Low High/Low High High/Low

A cursory look at the data reveals four incongruities:

<table>
<thead>
<tr>
<th>No</th>
<th>Pitch Registers</th>
<th>Lower Limits</th>
<th>Center Frequency</th>
<th>Upper Limits</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Extra low</td>
<td>71</td>
<td>80</td>
<td>88</td>
<td>17 Hz</td>
</tr>
<tr>
<td>2.</td>
<td>Low</td>
<td>89</td>
<td>100</td>
<td>113</td>
<td>24 Hz</td>
</tr>
<tr>
<td>3.</td>
<td>Mid</td>
<td>114</td>
<td>125</td>
<td>141</td>
<td>27 Hz</td>
</tr>
<tr>
<td>4.</td>
<td>High</td>
<td>142</td>
<td>160</td>
<td>176</td>
<td>34 Hz</td>
</tr>
<tr>
<td>5.</td>
<td>Extra high</td>
<td>177</td>
<td>200</td>
<td>225</td>
<td>48 Hz</td>
</tr>
</tbody>
</table>

Table 3: Critical Bands for Men

<table>
<thead>
<tr>
<th>No</th>
<th>Pitch Registers</th>
<th>Lower Limits</th>
<th>Center Frequency</th>
<th>Upper Limits</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Extra low</td>
<td>106</td>
<td>120</td>
<td>132</td>
<td>26 Hz</td>
</tr>
<tr>
<td>2.</td>
<td>Low</td>
<td>133</td>
<td>150</td>
<td>169</td>
<td>36 Hz</td>
</tr>
<tr>
<td>3.</td>
<td>Mid</td>
<td>170</td>
<td>185</td>
<td>211</td>
<td>42 Hz</td>
</tr>
<tr>
<td>4.</td>
<td>High</td>
<td>212</td>
<td>240</td>
<td>265</td>
<td>53 Hz</td>
</tr>
<tr>
<td>5.</td>
<td>Extra high</td>
<td>266</td>
<td>300</td>
<td>337</td>
<td>71 Hz</td>
</tr>
</tbody>
</table>

Table 4: Critical Bands for Women

12 It is ill-advised, according to Ladefoged (1968: xi), to base the acoustic phonetic accounts of a language on the speech of one or two informants. He contends that “this is a serious limitation; it may mean that any particular piece of data is really applicable to no more than a single speaker, and may not be typical of the language as a whole.” Nowadays, it is recommended that there be at least 20 participants: 10 males and 10 females for an acoustic phonetic study to be representative of a speech community.
1. The first has to do with the F0 of the first word [n] (the subject pronoun “I”) which is 105 Hz. Ahoua equates this F0 with a low register. Right next to the subject pronoun [n] is the verb [seli]. It is a disyllabic word. The F0 of the tone bearing unit [e] is 100 Hz. Ahoua equates this F0 with a high pitch register. How can an F0 of 105 Hz be low, while that of 100 Hz be high?

2. The second incongruity relates to the Tone Bearing Unit (TBU) of [i] in [seli] and the TBU of [ɛ] that immediately follows [i] of [seli]. The TBU of [i] is 92 Hz and is deemed to have a low register, whereas the TBU of [ɛ] in the [kɛ] that has a TBU of 70 Hz is classified as high. How can this be?

3. The third incongruity is found in the word [bla]. The nucleus is a bi-moraic, that is, the TBU [a] has two mora. The TBU of the first mora is 100 Hz and it is classified as high, while the TBU of the second mora, which is 135 Hz, is deemed to have a low pitch.

4. The fourth incongruity, the most puzzling of all, is stated as a general observation. How can the TBU of segment [kɛ], whose F0 is 70 Hz, be deemed high, while the F0 of the TBU of another segment whose frequency is 135 Hz be considered low?

To summarize, incongruities such as these defy mathematical logic. More importantly, these incongruities show that the interpretive framework used for accent languages does not work at all for tone languages. The reason is because accent and tone languages are processed and perceived differently. The former is processed in PT and perceived on a nominal scale, whereas the latter is processed in HG and perceived on a logarithmic scale.

5.1 Reanalysis of Ahoua’s Data in Light of CBT

When Ahoua’s example is reinterpreted using the CBT model, it yields excellent and expected results. Here is a CBT-based reanalysis of the correlations between F0 measurements and pitch registers:

| Sentence 1b: | [n  seli  kɛ  bla] |
| Indicative Mood: | I said that come (I told him/her to come) |
| F0 Measurements: | 105 100/92 70 100/135 |
| Pitch Register: | Low Low/Low Extra low Low/mid |

In 1997, Leben and Ahoua jointly wrote a paper in which they discussed what they considered to be a peculiar feature in Baule intonation. They coined the phrase “upstepping rule” to explain it. They found that the pitch registers on the TBUs of the last elements in the sentence keep rising. A CBT-based reanalysis of Sentence 1a as Sentence 1b above confirms the existence of this “upstepping rule.” We see that from the subordinating conjunction [kɛ] to the end of the sentence, there is a sustained rise in pitch from extra low, to low, to mid. This is one more example in addition to the ones provided in Koffi (2017) that show that CBT is very well suited to account for pitch registers in tone languages.

6.0 Summary

The examples above dramatize the need for acoustic phonetic frameworks that can account for the differential perception of pitch in accent and tone languages. The traditional prosodic distinction between accent and tone languages is grounded in neuroanatomical reality. What is
needed is a model for accounting for the differences and similarities in pitch perception between accent and tone languages in a principled way. The operations of this model can be summarized as follows. The acoustic signals emitted by the mouth of the talkers make their way into the auditory system where they are transformed into frequencies. However, the brain does not perceive these frequencies directly. They are first converted into nerve impulses that are transported through a complex neural transmission network to the thalamus. From there, electrochemical signals are sent to pitch perception centers. All pitch data, regardless of whether the language is accentual or tonal, go through HG for processing. If the language is tonal, all the processing takes place in HG and F0 information is correlated directly with certain pitch registers and interpreted on a logarithmic scale consisting of five registers: extra low, low, mid, high, and extra high. If the language is accentual, pitch data is ferried into PT for additional processing. Once there, autocorrelational analyses are quickly performed. The calculations lead to the simplification of pitch registers from five to two, whereby pitch is processed on a nominal scale either as low or high. The whole process is captured succinctly by the information displayed in Table 4.

<table>
<thead>
<tr>
<th>Language Type</th>
<th>Articulatory Production</th>
<th>Auditory Processing</th>
<th>Auditory Pathway</th>
<th>Cerebral Pitch Centers</th>
<th>Autocorrelation &amp; Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accent</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Planum Temporale</td>
<td>Nominal/categorical</td>
</tr>
<tr>
<td>Tone</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Heschl’s Gyrus</td>
<td>Logarithmic/fine-grained</td>
</tr>
</tbody>
</table>

Table 4: Pitch Perception and Processing Similarities and Difference

Licklider proposed DMPP in 1951 at a time when the understanding of the neural circuitry in the cortical system was still in its infancy. Advances in tonotopic mapping are now showing that DMPP accounts for pitch perception in accent languages very well. A consensus has emerged and is strengthened as tonotopic evidence accumulates and as HG and PT are found to be the two main pitch centers in the brain. Yet, questions remain and controversies abound. Plack (2004:13) writes that “Pitch has fascinated hearing researchers for hundreds of years, but despite all the interest, the nature of the auditory mechanisms that underlie our perceptions is still a matter of intense debate.” Zheng (2009:3079) adds that “Despite the theoretical importance of understanding the roles that PT regions play in the auditory cortical processing, functional attributes of these regions still remain to be fully characterized.” Schneider et al. (2005:1245) have ignited a controversy by claiming that the HG in the right hemisphere plays a greater role in pitch perception than previously expected. This finding is controversial because, if confirmed, it flies in the face of the long-established view that language functions are mostly lateralized in the left hemisphere unless pitch processing in the right hemisphere has to do with paralinguistic functions of F0. Bendor (2012:745) sums up his views on the matter as follows: “In conclusion, the existence of a pitch center in lateral Heschl’s gyrus remains a contentious issue among auditory neuroscientists.” These caveats notwithstanding, the tonotopic evidence and autocorrelational operations described in this paper highlight important neural processing differences between accent and tone languages which translate into perceptual differences. This means that two different analytical frameworks should be employed in describing pitch phenomena in these two

13 The DMPP would still work even if it were found that speakers of accent languages process pitch in the right hemisphere in HG. In that case, one would surmise that distance/delays in signal process would be responsible for pitch being perceived either as low or high. Signal degradation would be expected as neurons transmit pitch signals from the right hemisphere to the left hemisphere for processing.
types of languages. The model according to which a JND of only 1 Hz is all one needs to perceive one syllable as strong and another as weak works well for accent languages. However, the CBT-based model is better equipped to account for correlating F0 measurements with pitch registers in tone languages.

ABOUT THE AUTHORS

Ettien Koffi, Ph.D., is a professor of Linguistics at St. Cloud State University, Minnesota, USA, specializing in acoustic phonetics (Speech Intelligibility). His research interests center around sociophonetic variations in Central Minnesota English, acoustic phonetic accounts of intelligibility in L2 English, and acoustic phonetic and general description of Anyi, a West African Language spoken in Cote d'Ivoire. He is the author of four books and numerous papers covering topics as varied as syntax, translation, language planning and policy, orthography, and indigenous literacy training manuals. He can be reached at enkoffi@stcloudstate.edu.

References


