In Ida Toivonen, Piroska Cs´uri, Emile van der Zee , eds. *Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff*. MIT Press, in press.

Neural substrates for linguistic and musical abilities: a neurolinguist's perspective[1]

Yosef Grodzinsky

Departments of Linguistics and Neurology/Neurosurgery

McGill University

and

Edmond and Lily Safra Center for Brain Sciences

Language, Logic and Cognition Center

Department of Cognitive Science

The Hebrew University of Jerusalem

To many, music and language are hallmarks of humanity. Moreover, they seem to share formal properties: musical and linguistic expressions appear to be complex and rule governed, and their meaning seems compositionally determined. These apparent commonalities are among the reasons that have led some to wonder whether the abilities to produce and perceptually analyze linguistic and musical expressions are formally similar, and whether they have a common neural substrate. In this paper, I consider what would be required to demonstrate that language and music are cognitively and neurologically separable, proceed to a critical review the relevant literature, and propose a new experimental paradigm for the study of pitch accent, an element that has played an important role in recent studies. The basic idea is to replace the prevalent pitch discrimination paradigm with one based on sentence verification, in which semantic focus (marked by pitch accent), and *only* that associates to it, are used as a research tool.

It is tempting to say that musical and linguistic abilities, likely among the hallmarks of humanity, are similar. What comes to mind are not only formal properties and processing routines that these two abilities may share, but also, common brain mechanisms. In this chapter, I consider the logic of inquiry and the current state of empirical evidence as they pertain to the quest for common neural bases for language and music. I first try to enumerate the properties that any cognitive ability akin to language should possess (section 1), and move to a brief consideration of the neurological argument for the modularity of language from music (section 2). I then proceed to a critical review of studies that have investigated gross double dissociations between music/language (section 3). In section 4, I focus on studies of pitch discrimination in amusia, which I critique (section 5). In section 6, I propose a novel experimental paradigm for the study of pitch in language. Sensitive to past criticisms, I show that this paradigm escapes them. The paradigm, which I present in detail, is based on semantic considerations, specifically on the claim that *only* associates with focus (expressed via pitch accent). When an element in a sentence is focused, a set of alternative meanings emerges; *only* is a function that picks certain alternatives out of the focus set, and negates them. This paradigm helps to create minimal sentence pairs that need not be compared in order to test sensitivity to pitch accent. Rather, they can be investigated separately. This property of the materials helps the new paradigm get around criticisms proposed in the literature by Patel and his colleagues. I conclude (section 7) by alluding to salient properties of the speech of a famous amusical individual.

It is most pleasing to use this space for a discussion of focus in the context of music/language modularity, as these are two areas of inquiry to which Ray Jackendoff – an early teacher/mentor of mine – has made multiple, most valuable, contributions throughout his rich career (e.g., 1972; 1983, *passim*).

1. Human abilities akin to the linguistic

How can we tell that two (or more) classes of behaviors belong in the same cognitive unit? We must ask whether they are governed by the same set of building blocks and rules that combine them, structural constraints on such combinations, and algorithms that implement them in use. Osherson (1981) puts it very succinctly:

> … let C_1 and C_2 be two classes of processes and structures that conform to two sets of interlocking and explanatory principles, P_1 and P_2, respectively. If the properties of C_1 can be proved not to be deducible from P_2, and likewise for C_2 and P_1, then distinct faculties are (provisionally) revealed.

Fodor (1983) suggests several perspectives from which the modularity of cognitive systems from one another can be assessed: **a**. the computational perspective, in which we inquire whether the structural principles (*a k a* knowledge) that govern one system can be deduced from those of another; **b**. the implementational perspective, which examines identity or distinctness of the processes that implement this knowledge in use. **c**. the developmental perspective, which looks at similarities and differences in the way cognitive systems unfold in the developing child; **d**. the neurological perspective, which explores anatomical and physiological properties brain loci that support each system.

Given what we currently know about language, here are some properties we should require from a neurocognitive ability akin to language:

I. It must be able to handle (i.e., analyze, perhaps even produce) strings, or continua, that unfold over time.

II. It must be able to concatenate smaller forms into bigger ones by combinatorial rules, to ensure rich expressiveness. These must be constrained by principles similar to linguistic ones.

III. Its inventory of basic forms must be meaning bearing, where smaller pieces of meaning compose into larger ones.

IV. Diversity of forms and rules is permissible, as long as it is constrained by universal principles.

V. Its dedicated mechanisms must be supported by specialized neural clusters.

This list helps us home in on three suspects: mathematical and musical abilities, and the ability to sequence motor actions. Each of these seems to be a serious candidate to satisfying criteria I-IV. Indeed, some have maintained that language and mathematics share a common cognitive basis (Changeux, Connes and DeBevoise 1998; Chomsky 1988; Henschen 1920), while others have argued that the same holds for language and motor ability (Schuell 1965; Kimura 1973a,b; Rizzolatti and Arbib 1998; Fadiga, Craighero and Roy 2006).

Neurologically, we know that each of these abilities is associated with a disorder or deficit, which may lead to the satisfaction of criterion V, the focus of this chapter:

- Linguistic ability   aphasia
- Mathematical ability   acalculia
- Motor ability   apraxia
- Musical ability   amusia

Aphasia, acalculia, and apraxia manifest subsequent to focal brain damage, which leads to debates regarding criterion V. Recent experimental evidence has suggested that language and mathematics are neurologically separable (Brannon 2005; Cohen and Dehaene 2000; Gelman and Butterworth 2005).[2] Regarding motor abilities, there, too, have been claims for and against modularity, most notably in the context of the Mirror Neuron theory (e.g., Rizzolatti and Arbib 1998; Pülvermüller and Fadiga 2010; Fazio et al. 2009; see Grodzinsky 2006, 2013; Venezia and Hickok 2009 for critical approaches).

In the case of music, matters are more complicated. Not only is neurological evidence scarce, but also, discussions of differences and similarities between language and music have been rather complex.

In his famous Norton Lectures, Leonard Bernstein (1973) proposed to try to find true parallels between language and music, since "all musical thinkers agree that there is such a thing as a musical syntax, comparable to a descriptive syntax of speech" (lecture 2).[3] In the same spirit, Lehrdahl and Jackendoff (1980) claimed to have found deep parallels between language and music. Jackendoff

(2009) has further asked whether there is domain specificity for language: "What does music share with language that makes them distinct from other human activities?" (195). Katz and Pesetsky (2011) have gone even further, formulating the *Identity Thesis for Language and Music*: "All formal differences between language and music are a consequence of differences in their fundamental building blocks (arbitrary pairings of sound and meaning in the case of language; pitch-classes and pitch-class combinations in the case of music). In all other respects, language and music are identical."

Bernstein, Jackendoff and Lehrdahl and Katz and Pesetsky discuss the relation between language and music from a representation and operational (processing) perspective. They steer clear from the neurological perspective, on which I will henceforth focus. The question here, then, will be: are there common neural substrates for musical and linguistic processes? At present, the evidence doesn't tell us as much as we'd like it to. We can nonetheless try to think about new ways to explore it, but that is not easy. In what follows, I will look at the form of the neurological argument in each of its incarnations, and try to see what conclusions, if any, can be drawn from the evidence at hand.

2. The neurological argument for the separability of language from music

Schematically, tests of neurological modularity have the structure in (1):

(1)	***Functional anatomy type 1***	***Functional anatomy type 2***
Measured variable A	+	–
Measured variable B	–	+

That is, to demonstrate neurological modularity and in keeping with Osherson's dictum, tests that measure variables *A*, *B*, must produce different values in neurologically distinct areas of functional types *1*, *2*. The putative result in (1) would therefore indicate that the neural basis of the cognitive component(s) probed by test *A* are supported by area(s), or functional type, *1*, whereas *B* is supported by *2*. Crucially, *A*, *B*, are distinct. This is the well-known argument from *Double-Dissociation* (*DD* henceforth).

The *DD* argument can be applied in several ways, as the rows and columns in (1) can have different headers, as detailed in (2):

(2) ***Pieces of the neurological argument***

a. Types of functional anatomy: loci of lesion; loci of activation clusters
b. Types of measured (dependent) variables: (i) behavior along some dimension, (ii) brain activity due to behavioral manipulation
c. The behavior pieces chosen (driven by a cognitive theory)

In what follows, I will offer a critical review of past work along the lines detailed in (1)-(2). I will then follow with a constructive proposal.

3. Gross Double Dissociations in disease and in health

Traditionally, neuropsychologists have been engaged in a search for *DD*s. In the present context, the quest has been for cases in which language is severely disrupted whereas music remained intact, juxtaposed to cases in which language is intact, but music is gone, as schematized in (3), where the measured variables *A*, *B*, of table (1) are replaced by non-specific tests of musical and linguistic ability, and the functional areas are replaced by missing (lesioned) brain regions:

(3)	***Lesion in brain locus 1***	***Lesion in brain locus 2***
Musical ability	High performance	Low performance
Language ability	Low performance	High performance

Plainly put, the expectation here is to observe aphasia without amusia and vice versa. Such cases seem to exist (Peretz 1993; Peretz and Coltheart 2003; Grodzinsky and Finkel 1998), and their functional impairment is described as follows:

G.L. (Peretz 1993) is a Québec man who apparently has amusia without aphasia. He has “lesions in the both right and left superior temporal gyri, temporal poles, inferior frontal gyri and insulae.” “Out of 140 musical excerpts…familiar to everyone in Québec …he could not identify a single one… he was able to discriminate changes between single pitches…was sensitive to differences in melodic contour in short melodies. Yet he showed an absence of sensitivity to musical key.” Language was largely intact. He scored 32/36 on the Token Test. “He scored in the normal range on standardized aphasia tests.”

J.C. (Grodzinsky and Finkel 1998) is a woman who apparently suffers from aphasia without amusia. She has “a fronto-temporal lesion, including Broca’s area,” her speech is non-fluent and agrammatic, she speaks in short utterances and omits functional vocabulary. Her musical abilities are intact: an opera singer and a voice teacher prior to the cerebro-vascular incident that impaired her, she can still sing rather well.[4]

(4)	***Patient G.L.:*** ***Lesion excludes Broca’s region***	***Patient J.C.:*** ***Lesion includes Broca’s region***
music	Low performance	High performance
language	High performance	Low performance

J.C.'s comprehension deficit was documented in detail (Grodzinsky and Finkel 1998). It consisted of a syntactic impairment, manifested through deficient performance on a forced-binary-choice Sentence-to-Picture Matching task (5), and on a Grammaticality Judgment task (6) that featured grammatical sentences (6a,c) as well as violations (6b,d). The symbol "◄" represents the extraction site:

(5) ***Comprehension performance***	***% correct***
a. The woman who dried the girl was thin	*70*
b. The woman who the girl dried ◄ was thin	*40*

(6) ***Judgment of well-formedness***	***% correct***
a. It seems to Sally that the father rewards himself	*80*
b. *It seems to Sally that the father rewards herself	*70*
c. The father seems to Sally to ◄ reward himself	*40*
d. *The father seems to Sally ◄ to insult herself	*30*

The documented performance of these patients suggests that aspects of language and music are indeed doubly dissociated.

A similar logic has guided inquiries with healthy populations – the idea has been to search for double dissociations in the healthy brain. For example, in fMRI, a linguistic task is expected to activate neuronal aggregate *X* but not *Y*; whereas a musical task would activate cell aggregate *Y* but not *X*. Thus Koelsch (2005) reports ERP and fMRI studies of well formedness, in which activations of ±well-formed sentences were compared to regular and irregular (tonal) musical pieces. This schematic design is in (7), where Δ represents the difference in activation level between test and control (i.e., between brain activation with +well-formed continua stimuli and –well-formed ones):

(7)	***Brain loci for language***	***Brain loci for music***
Music *(±well-formed)*	Low activation Δ	High activation Δ
Language *(±well-formed)*	High activation Δ	Low activation Δ

Koelsch reports large bi-lateral frontal and temporal (perhaps temporo-parietal) regions that are activated by the musical contrast, which he juxtaposes to left Brodmann Areas 44,45 – regions traditionally thought to be activated by syntax.

Using the same logic, Fedorenko et al. (2012) have monitored the fMRI signal with a different set of contrasts. That is, they used ±scrambled sentences, which they compared to songs. Their study had the following schematic design:

(8)	***Brain loci for language***	***Brain loci for music***
Music *(±Scrambled)*	Low activation Δ	High activation Δ
Language (*±Scrambled****)***	High activation Δ	Low activation Δ

Loci for language were found in the left inferior frontal gyrus (LIFG, roughly Broca's region), left middle frontal gyrus, and left anterior, middle and posterior temporal regions, as well as the angular gyrus.[5] Music areas were found on both hemispheres, from right and left anterior and posterior temporal regions, to right and left premotor, supplementary motor, areas. Again, a double dissociation is demonstrated, but not as sharply as one would have wished.

We might examine the results of these studies – whether they evince neuroanatomical overlap between language and music; and whether we observe a match between the lesion studies and those in health, or even anatomical congruence between the 2 sets of fMRI studies. However, before looking at the results, we might question the choice of tasks, materials, and contrasts:

I. *Are the musical and linguistic materials and contrasts uniform?* If different studies use different types of contrasts, why would one expect the resulting errors (in the case of lesion work) or activation patterns (in health) to be similar in the first place?

II. *Are the musical and linguistic tasks matched?* The Specificity/modularity agenda requires use of parallel methodology and reasoning across cognitive domains.

III. *How do the tests connect to linguistic and musical structure?* The interest in the relation between language and music stems from the belief that linguistic and musical strings are structured and governed by rules. We also know that the neuropsychology of other domains indicates complex symptomatology that differentiates between different syndrome types *within* each domain. How does this structural complexity enter into the considerations here?

Reviewing the studies above, we begin with the neuropsychological cases: G.L. and J.C. received a mixed bag of tests. G.L.'s linguistic abilities were assessed through the Token Test (de Renzi and Vignolo 1962). However it is not clear that this test assesses linguistic, as opposed to general cognitive, skills: it presents a display of shapes in different colors and sizes, and requests the subject to act on statements whose "complexity" (on a metric that has little to do with linguistic structure) is varied. E.g., in a context of large objects only – where the properties to be attended to are shape and color – the command may be "*pick up the yellow circle*"; in a context that includes all objects, with more properties to attend to than before (shape, color, size), the command may be "*pick up the small yellow rectangle AND the large red circle.*" With this structure, it is very difficult ascertain that the dimension on which difficulty increases here is linguistic. The increased length of command, and its appeal to a larger number of spatial properties of shapes, might well tap some general cognitive resource in an incremental fashion. In other words, G.L.'s success on the Token Test is not indicative of full linguistic ability, as this test may well have missed fine linguistic deficits.

Musically, G.L. was asked (and failed) to identify tunes that were well-known in his culture at the time. There was nothing in this test to suggest a direct analogy or parallel to the linguistic test just described.

J.C. was given a very different language test battery, in which syntactic structure was varied systematically along the Movement dimension, and constraints on syntactic movement were occasionally violated. In the musical domain, she was not tested formally, but as she actively sang, we are fortunate to have access to recordings of her singing ability. Her singing, good as it was, did not necessarily tap all her musical abilities; in particular, it was not designed to match the linguistic materials in terms of structural complexity or difficulty. Thus on all counts, G.L. and J.C. do not constitute a double dissociation. The road to such dissociations appears long and treacherous.

More recent studies of music/language relations in health appear to be finer grained in terms of the choices made. Still, we might want to scrutinize the materials and contrasts chosen, as well as the tasks. We should also review the degree of cross modal matching - the extent to which tests carried out in different modalities are matched in terms of the generic resources they require. One useful review is Koelsch (2005), who looked at studies in which violation of musical expectation in musical continua were compared to violations of grammaticality in language. Regrettably, there is no discussion of the nature of the violations in both domains and the rationale behinds the choices made; nor is a parallelism between the violations across domains established. One wonders, therefore, whether the contrasts chosen are (i) representative of their respective domains in a theoretically justifiable fashion, and (ii) whether they are parallel.

Fedorenko et al. (2012) carried out an fMRI which used a collection of English sentences and Western musical pieces. Both sentences and musical pieces were presumably matched, and were further contrasted to their scrambled versions (i.e., scrambled sentences and songs). Their participants were asked, in the music task, *"how much did you like this piece?"*, and in the language task: *"did X feature in the stimulus?"*, where *X* was a given 'memory probe.' The authors search for Regions Of Interest (ROIs) in the brain, defining them on the basis of a functionally selective activation pattern they exhibit (i.e., in terms of the Δ between brain responses to scrambled vs. non-scrambled stimuli was low for music and high for language). They are therefore known as fROIs (as opposed to anatomical ROIs, defined by anatomical properties such as borders or topography). Fedorenko et al.'s first goal was to identify fROIs which are sensitive to language but not music (fROI-1), as well as opposite functional regions (fROI-2). Finding such a *DD*, they argue for a functional double dissociation (3294).

But in light of the above discussion, one might wonder about the *motivation* for the choice of basic stimuli, whether the language and musical *tasks were on a par*, and moreover whether *scrambling* the stimuli is a theoretically interpretable manipulation. Finally, it is not clear that the tasks in the two domains are parallel. The absence of discussion of any of these issues leaves a reader puzzled. These authors' justification of the choice of linguistic and musical continua comes from the fact that all these continua activate clusters in each individual subject – an interesting observation, but hardly a key to interpretation. As well, there is no reason to think that the tasks were on a par – one requesting a "likeability" ranking, another – word monitoring. These choices, as well as the principles that may underlie the irregularity induced by scrambling in each domain, are not discussed any further. The materials and tasks are thus left as a black box, which seems to preclude a conclusion of the sort that Fedorenko et al. wish to draw. It would appear, then, that theoretically motivated and better matched tests would be needed in order to evaluate whether language and music are supported by the same brain regions.

Let me be a bit more specific. One would imagine that Fedorenko et al.'s interest in differences and commonalities between language and music stems from the fact that sentences and musical pieces are structured, rule-governed objects. Indeed, they seem to suppose that any task that involves musical combinations (compared to blatant violations of combinatorial rules) is comparable to a task that involves linguistic combinations (compared to blatant violations of combinatorial rules). By this logic, the tasks they used were on a par.

But differences and commonalities between these two presumed "faculties" or "modules" can only be established through a detailed and precise specification of the combinatorial rules at issue. Only this way can a valid comparison be established, and thus ignoring the details (as Fedorenko et al. do in this case) does not really help. In order to argue that music and language are distinct, we need to ascertain that similar musical and linguistic contrasts and tasks indeed tapped different neural resources. An argument for the modularity of these two faculties would first require a demonstration that the contrasts used were equally taxing – that task demands tapped the same structural principles/combinatorial rules. This is not likely to have been the case here, and at any rate, no discussion of this issue is found in the paper. As a result, we are left in the dark.

The foregoing discussion and critique leads to several desiderata from a proper design for music/language experiments. To be truly informative, such experiments should:

- Make an explicit connection to theories of musical and linguistic knowledge
- Keep task demands parallel across modalities and groups
- Focus on cognitive dimensions that are relevant to structural analysis in all domains

Once the "right cognitive dimension" is found, the *DD* schema in (1) can be refined. Below are sketchy design tables, for studies aimed to detect double dissociations in disease via selective performance deficiencies (9), and in health via localized signal intensity differences (10), which I develop below:

(9) *DD*s IN LESION STUDIES	***Functional Deficiency A***	***Functional Deficiency B***
Music *("right cognitive dimension")*	Low performance	High performance
Language *("right cognitive dimension")*	High performance	Low performance

(10) *DDs* IN fMRI SIGNAL DETECTION IN HEALTHY INDIVIDUALS	***Brain region A***	***Brain region B***
Music *(±"right cognitive dimension")*	Low activation Δ	High activation Δ
Language *(±"right cognitive dimension")*	High activation Δ	Low activation Δ

Next, I will try to illustrate how such a research program is implemented.

4. Pitch discrimination in amusia

Pitch is that quality of sound which allows us to play musical melodies. It moreover represents an abstraction: many different sounds have the same pitch (Schnupp et al. 2011, Chapter 3). It is an abstraction, as many different instruments (and voices) can produce sounds with the same pitch. It is thus among the most important properties of sound that help humans make music (Nelken 2011). No wonder, then, that it has featured in the research program that attempts to connect music to the neural tissue that support it (E.g., Ayotte et al. 2002; Hyde et al. 2011; Patel 2012). It is also important for linguistic meaning and communication: e.g., the difference between a sentences in which a different element is focused as manifest by pitch accent (e.g., between *HE congratulated you* and *he congRATUlated you*).

Pitch therefore avails us of a possible dimension along which we can compare the linguistic and the musical. Indeed, several studies have attempted to identify pitch discrimination problems in the linguistic context in so-called amusical individuals, who suffer musical pitch deficits. Ayotte et al. (2002) asked these individuals to detect differences in melody pairs that differed in 1 semi-tone (positioned quasi-randomly), and then presented them with a language task, in which they were asked to indicate whether 2 sentences are the same or different, where the pairs consisted of sentences that differ in pitch accent:

Sentences compared (Ayotte et al., 2002)
(11) a. Sing **NOW** please! b. **SING** now please!

Ayotte et al. found that amusical individuals were near normal in discriminating between these sentences. Their success here, contrasted with their failure on the musical discrimination task, led Ayotte et al. to conclude that music and language are modular from one another.

Patel et al. (2008) and Liu et al. (2010) disagree with this conclusion. To them, the high performance on (11) is not particularly telling. Amusicals may have succeeded on (11) because “salient pitch changes can be ‘tagged’ according to the syllable on which this occurs, thus reducing the memory demands of the task” (Liu et al. 2010, 1683). The idea is that location of the pitch rise in the sentence could thus serve as a cue in the comparison task. Patel and colleagues therefore suggest to ignore the Ayotte et al. (2002) result, and move on to instances in which “tagging” is not an option, like question/statement pairs (12)-(13), where the pitch difference was always in the same (sentence-final) position[6]. This linguistic material, they argue, would be a better test of language-music modularity, as the materials are now better matched. And so these materials were administered with 2 types of tasks – discrimination (same/different), and identification (question or statement):

Stimuli from Liu et al. (2010)
(12) a. She looks like Ann**!** b. She looks like Ann**?**
(13) a. He was born in Illinois**!** b. He was born in Illinois**?**

Indeed, those individuals who had serious trouble with the musical comparison task were also not good at distinguishing questions from statements (12)-(13). Patel and colleagues conclude that this result – the cross-modal cooccurrence of failures – argues *against* domain specificity, as the musical deficit co-occurs with a linguistic one.

Still, the conclusion reached by Patel and his colleagues may be a bit hasty. As the stakes are high – at issue is music/language modularity – I would like to revisit Ayotte et al.'s results for (11), and see whether a different interpretation is possible. I will then propose a way to get around the experimental problems noted by Patel and his colleagues, one that might lead to an improved test, with the hope of obtaining a somewhat higher resolution than previous studies.

5. A critique of the pitch discrimination studies

The situation as presented, then, is as follows:

(14)	***Amusical performance***
Music *(Δ pitch)*	*Low discrimination*
Language *a.* Focus location *(Δ pitch) - (11)*	*High discrimination*
b. Question/statement *(Δ pitch) - (12)-(13)*	*Low discrimination*

Patel and his colleagues reject the relevance of the focus discrimination test in (11), arguing that those in (12)-(13) are more informative. But there may be reasons to take the opposite view – to argue that in fact the amusicals' success on the contrast in (11) is a better benchmark of their pitch identification in the linguistic context than their failure in (12)-(13). In what follows, I will try to argue for the latter view. A successful argument would hopefully reopen the possibility for an empirical argument in favor of language/music modularity.

To begin with, let me note that amusical subjects, said to fail in recognition and imitation tasks with familiar musical pieces, reportedly have normal communicative skills.[7] And yet, if the failure of these individuals to discern a question from a statement (11)-(12) is indicative of a communication deficit, why is it not manifest in their daily linguistic functioning? It is true that many communicative acts contain many cues beyond pitch regarding semantic type, but there surely are instances in which such a discrimination deficit would manifest in communication. As Liu et al. (2010) point out, "amusics rarely report problems outside the musical domain," but proceed to suggest that "it may be expected that these individuals would struggle with aspects of spoken language that rely on pitch-varying information" (1682). Curiously, the amusicals' performance level on the imitation task, while lower than normal (only 87% correct), was much higher than their chance performance in identification or discrimination. Liu et al., while acknowledging the absence of noticeable communicative deficits in amusical subjects, nonetheless insist that "pitch deficits can be behaviorally relevant to both speech and music" (1691), offering no further discussion.

Next, consider the argument that linguistic pitch is carried by meaning- and form-bearing objects, whereas and musical pitch is not. Patel (2012) proposed this distinction to account for the amusicals' success in the focus discrimination condition (11). The idea is that pitch is linked to a word (perhaps to a syllable), whereas pitch contours without language are not, and that this link might have eased memory demands. Perhaps, but a question immediately arises: why are the same subjects worse with question/statement pairs, that also have syllabic, lexical, and propositional content? Moreover, do we understand the reasons behind the differential performance found between discrimination/identification and imitation of questions and statements?

These questions remain unanswered, leading to apparent inconsistencies in the data. In light of these, I would like to suggest ways to revisit the language/music modularity question in a manner that gets around some of the problems.

6. A proposal: focus structures with only and without

In this section, I will put forward a simple proposal for an improved pitch test in the linguistic domain one that would get around Patel's "tagging" critique of Ayotte et al.'s focus discrimination study, and would also be on a par with the typical musical recognition task that in which amusical subjects fail. The goal here is to situate the tasks in a more naturalistic context, that would not require a comparison (made easy by "tagging"), and moreover would not be taxing in a way that isn't necessarily relevant to communication. For that, I will propose a task in which pitch is required for linguistic (as opposed to meta-linguistic) analysis in sentence comprehension. That is, as amusical subjects fail in simple tasks in which pitch is crucial, namely recognition of familiar musical pieces (let alone singing them or detecting deviations from melodic lines), we might want to create a linguistic analogue, in which difference in pitch accent would be crucial for language use. Natural candidates tasks involve comprehension, question answering, or verification. I will focus on the latter, in the hope of finding a test for Patel's claim that the discrimination between different sites of pitch accent within a sentence is not a valid test of sensitivity to pitch.

In many of the world's language (though by no means all), semantic focus is triggered by pitch accent. Semantic focus evokes a set of alternatives, picks out one, and makes it more salient. Ray Jackendoff made an early contribution to the analysis of focus in the generative framework, analyzing it through the use of the "structured-meaning" approach; later, alternative semantics was introduced (Rooth 1985; 1992), which is what guides my brief presentation below. In (15a), we have a sentence *p*, and focus on an element within *p* evokes a set of alternatives whose members are all propositions that John introduced Bill to someone in the context.

Simplifying somewhat, assume a context C that features a scenario in which John (and only he) is introducing people to one another, and where the other participants are Bill, Mary, Betty and Sue. Focus on Sue, conveyed through pitch accent, asserts the proposition *p* (15b), and in addition gives rise to additional, focus semantic value by allowing a set of alternatives A_c^{SUE} (15c-d), of which one is made more salient by focus:

(15)	***Focus evokes an alternative set A_C***
	a. John introduced Bill to SUE *(though there were others present)* b. *p = John introduced Bill to Sue* c. A_c^{SUE} = {x∈D_e/John introduced Bill to x}= ={*John introduced Bill to Sue*, *John introduced Bill to Betty*, *John introduced Bill to Mary*}[8]

Focus, then, underscores the meaning in which the alternative containing *p*, *John introduced Bill to Sue*, is made the most salient one. We need not get into the details of the mechanism here. Suffice it to note that the critical element for us here is the set of alternatives that focus gives rise to, and that this set varies with the focused element. Thus, in (16) below, the set of alternatives A_C' is different from that in (15) above:

(16)	***Focus evokes an alternative set $A_{C'}$***
	a. John introduced BILL to Sue *(though there were others present)* b. *p = John introduced Bill to Sue* c. $A_{c'}^{\mathbf{BILL}}$ = {x∈D_e/John introduced x to Sue}= ={*John introduced Bill to Sue*, *John introduced Betty to Sue*, *John introduced Mary to Sue*}

Sentences (15a) and (16a) make the same assertion *p*, but differ in their focus semantic value, as $A_c^{\mathbf{SUE}} \neq A_{c'}^{\mathbf{BILL}}$ – different sets of alternatives, evoked when the pitch accent is on Bill or Sue. Thus a scenario in which John introduced Bill to someone, who happened to be Sue, is compatible with (15) but not (16), whereas a situation in which John introduced someone to Sue, and that someone was Bill, is compatible with (16) but not (15). The respective acceptability judgments follow.

The meaning differences between (15) and (16) may seem somewhat murky, because focus makes a certain alternative more salient than others, and the notion of salience is somewhat difficult to capture. However, matters become crystal clear when *only* is introduced as an element that associates with focus (Rooth 1985). Sentential *only* is a function that combines with a sentence *p* and a set of alternatives A_c that focus evokes (i.e., the set of all non-weaker alternative propositions to *p* that is supplied by C), and returns a set of propositions in which all but *p* are negated. The result is a sentence that asserts *p*, where all the other alternatives in A_c are false (Rooth 1985; Fox 2007):

(17)	[[*only*]] ($A_{<st,t>}$)(p_{st}) = λw: *p*(w) =1.$\forall q \in$ A: *q*(w)=0 [9]
	a. Informally: *only* in a function that takes a proposition *p*, a world w and a set of alternatives A, presupposes that *p* is true in w, and makes false in w every proposition *q* which is non-weaker than *p*.

A concrete application is given in (18):

(18)	John **only** introduced Bill to **SUE** *(and to no other individual present)*
	a. *p=John introduced Bill to Sue*
	b. A_c^{SUE}={x∈D_e/John introduced Bill to x}={*John introduced Bill to Sue*, *John introduced Bill to Betty*, *John introduced Bill to Mary*}
	c. *Only*(*p*)(A_c^{SUE})={*John introduced Bill to Sue*, ¬(*John introduced Bill to Betty)*, ¬(*John introduced Bill to Mary)*}
	d. It is true that that John introduced Bill to Sue, but it is false that John introduced Bill to Betty, and it is false that John introduced Bill to Mary

When *only* associates with another focused element in *p*, the result is a different meaning, because $A_{c'} \neq A_c$, and the application of *only* to it would negate different alternatives:

(19)	John **only** introduced **BILL** to Sue *(and to no other individual present)*
	a. *p=John introduced Bill to Sue*
	b. $A_{c'}^{\mathbf{BILL}}$ = {x∈D_e/John introduced x to Sue}={*John introduced Bill to Sue, John introduced Betty to Sue, John introduced Mary to Sue*}
	c. *Only*(*p*)($A_{c'}^{\mathbf{BILL}}$) ={*John introduced Bill to Sue, ¬(John introduced Betty to Sue), ¬(John introduced Mary to Sue)*}
	d. It is true that John introduced Bill to Sue, but it is false that John introduced Betty to Sue, and it is false that John introduced Mary to Sue

We can now see that although (18) and (19) make the same assertion *p*, they have different truth-conditions, because pitch accent marks a different element in each case, thereby evoking a different set of alternatives. *Only* then negates every proposition *q≠p*:

(20)	*Alternative sets of (18) vs. (19)*
	a. $A_c^{\mathbf{SUE}}$ ={x∈D_e/John introduced Bill to x}={*John introduced Bill to Sue*, *John introduced Bill to Betty*, *John introduced Bill to Mary*}
	b. $A_{c'}^{\mathbf{BILL}}$ = {x∈D_e/John introduced x to Sue}={*John introduced Bill to Sue, John introduced Betty to Sue, John introduced Mary to Sue*}

The reader may have noticed that the examples chosen above all involve a ditransitive predicate (*introduce*). This is done on purpose, in order to make the meaning contrast that different focus choices produce as minimal as possible. The idea here is to create a task whose performance requires sensitivity to pitch accent, and where pitch accent is placed on elements that are syntactically and semantically on a par, modulo the task at hand.[10] *Only* needs focus, and our task would include a possible position for association with focus on each of the 2 objects of the ditransitive verb. Normal performance on a verification task, given a scenario, would require the identification of focus location, which would occur in the absence of a comparison between 2 representations. In this task, “tagging,” as postulated by Patel and his colleagues, is not possible.

Let me provide a concrete example of how this meaning contrast is produced:

(21)	_Scenario C_: John made several introductions. He introduced Bill to Sue. He then introduced Mary to Betty. Finally, he Mary introduced to Sue. There were no other introductions.		
	Sentences:	a. John **only** introduced Bill to **SUE**	True in C
		b. John **only** introduced **BILL** to Sue	False in C

Let’s analyze what happened in each case. Scenario (context) C makes the assertion in both (21a) and (21b) true. Yet, C makes no member of the alternative set of (21a) false, which is hence true. However, (21b) contains the proposition ¬*(John introduced Mary to Sue)*, which is false in C, as the reader may verify.

The reader may likewise verify that scenario C', described in (22), produces opposite results:

(22)	*Scenario C'*: John made several introductions. He introduced Bill to Mary. He then introduced Mary to Betty. Finally, he introduced Bill to Sue. There were no other introductions.		
	Sentences:	a. John **only** introduced Bill to **SUE**	False in C'
		b. John **only** introduced **BILL** to Sue	True in C'

The above sketch makes it quite clear, I hope, that this setup – the association of *only* with focus – allows for the testing of sensitivity to pitch accent in a task that does not require discrimination. When the right controls are introduced (and there are many, to be sure), this should allow for testing through a verification (truth-value judgment) task. It is equally easy to imagine, I think, a production task with scenarios like (21), (22), in which amusical subjects would be forced to use *only*, and the issue would be whether or not they can successfully use pitch accent to mark the associated focus.

An implementation of this proposal is presently unavailable. What is important about it is that the above does not enable "tagging," because no comparison or discrimination between two utterances is required. Patel et al.

would predict that amusical subjects would fail in this verification task. Failure on their part would provide strong empirical evidence against the modularity of language and music. And thus, while at present, no relevant result is available, the jury appears to be still out on the modularity of language and music, at least until a result of the proposed experiment, or some related one, is obtained.

7. Coda

I tried to revive the notion that amusia, as reported in the clinical literature, does not co-occur with a language deficit (contra Liu et al. 2010). One anectodal, yet not insignificant, observation relates to the famous late economist Milton Friedman, believed to be amusical. His fame allows us to have access to speech samples of his. An important example is an interview on Greed he granted Phil Donahue in 1979.[11] If you haven't seen it, I would urge you to do so, for Friedman's especially expressive intonation, containing many questions and exclamations (apparently intended to make his argumentation more convincing) might make a compelling case for language/music modularity.

REFERENCES:

Ayotte, Julie, Isabelle Peretz and Krista Hyde. 2002. Congenital amusia. *Brain* 125: 238-251.

Bernstein, Leonard. 1973. *The Unanswered Question*. Cambridge, MA: Harvard University Press.

Brannon Elizabeth M. 2005. The independence of language and mathematical reasoning. *Proceedings of the National Academy of Sciences of the United States of America* 102: 3177-3178.

Changeux, Jean-Pierre, Alain Connes, and M.B. DeBevoise. 1998. *Conversations on Mind, Matter, and Mathematics*. Princeton, NJ: Princeton University Press.

Chomsky, Noam. 1988. *Language and Problems of Knowledge: The Managua Lectures.* New York: Cambridge University Press.

Cohen, Laurent and Stanislas Dehaene. 2000. Calculating without reading: unsuspected residual abilities in pure alexia. *Cognitive Neuropsychology* 17: 563-583.

Deschamps, Isabelle, Galit Agmon, Yonatan Loewenstein and Yosef Grodzinsky. Submitted. Quantities and Quantifiers: Weber's Law, Monotonicity and Modularity. Ms. McGill and HUJI.

Fadiga, Luciano, Laila Craighero, and Alice Roy. 2006. Broca's Region: A Speech Area? In *Broca's Region*, edited by Yosef Grodzinsky and Karin Amunts, 137-152. New York: Oxford University Press.

Fazio, Patrik, Anna Cantagallo, Laila Craighero, Alessandro D'Ausilio, Alice C. Roy, Thierry Pozzo, Ferdinando Calzolari, Enrico Granieri, and Luciano Fadiga. 2009. Encoding of human action in Broca's area. *Brain* 132: 1980-1988.

Fedorenko, Evelina, Josh McDermott and Nancy Kanwisher. 2012. Sensitivity to musical structure in the human brain. *Journal of Neurophysiology* 108: 3289-3300.

Fodor, Jerry. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fox, Danny. 2007. Free Choice Disjunction and the Theory of Scalar Implicatures. In *Presupposition and Implicature in Compositional Semantics*, edited by Uli Sauerland and Penka Stateva, 71-120. New York: Palgrave Macmillan.

Gelman, Rochel and Brian Butterworth. 2005. Number and language: how are they related? *Trends in Cognitive sciences* 9: 6-10.

Grodzinsky, Yosef. 2006. The language faculty, Broca's region, and the mirror system. *Cortex* 42: 464-468.

Grodzinsky, Yosef. 2013. The mirror theory of language: a neuro-linguist's perspective. In *Language Down the Garden Path: The Cognitive and Biological Basis for Linguistic Structure*, edited by Montserrat Sanz, Itziar

Laka, and Michael Tanenhaus, 333-47. Oxford, UK: Oxford University Press.
Grodzinsky, Yosef and Lisa Finkel. 1998. The neurology of empty categories. *Journal of Cognitive Neuroscience* 10: 281-292.
Henschen SE. 1920. *Klinische und anatomische Beitraege zur Pathologie des Gehirns.* Stockholm: Nordiska Bokhandeln.
Heim, Stefan, Katrin Amunts, Dan Drai, Simon Eickhoff, Sara Hautvast and Yosef Grodzinsky. 2012. The language-number interface in the brain: A complex parametric study of quantifiers and quantities. *Frontiers in Evolutionary Neuroscience* 4 (4): 1-12.
Hyde, Krista L., Robert J. Zatorre and Isabelle Peretz. 2011. Functional MRI Evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex* 21: 292-299
Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar.* Cambridge, MA: MIT Press.
Jackendoff, Ray. 1983. *Semantics and Cognition.* Cambridge, MA: MIT Press.
Jackendoff, Ray. 2009. Parallels and Non-Parallels between Language and Music. *Music Perception* 26: 195-204.
Jentschke, Sebastian and Stefan Koelsch. 2008. Musical training modulates the development of syntax processing in children. *NeuroImage* 47: 735-744.
Katz, Jonah and David Pesetsky. 2011. The identity thesis for language and music. Ms. Institute Jean Nicod and MIT.
Kimura, Doreen. 1973a. Manual activity during speaking— I. Right-handers. *Neuropsychologia* 11: 45-50.
Kimura, Doreen. 1973b. Manual activity during speaking— II. Left-handers. *Neuropsychologia* 11: 51-55.
Koelsch, Stefan. 2005. Neural substrates of processing syntax and semantics in music. *Current Opinion in Biology* 18: 207-212.
Lehrdahl, Fred and Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
Liu, Fang, Aniruddh Patel, Adrian Fourchin and Lauren Stewart. 2010. Intonational processing in congenital amusia: discrimination, identification and imitation. *Brain* 133: 1682-1693.
Nelken, Israel. 2011. Music and the auditory brain: where is the connection? *Frontiers in Human Neuroscience* 5: 106.
Osherson, Daniel. 1981. Modularity as an issue for cognitive science. *Cognition* 10: 231-232.
Patel, Aniruddh. 2012. Language, music, and the brain: a resource-sharing framework. In *Language and Music as Cognitive Systems*, edited by Patrick Rebuschat, Martin Rohrmeier, John A. Hawkins, and Ian Cross, 204-223. Oxford: Oxford University Press.

Patel, Aniruddh, Meredith Wong, Jessica Foxton, Aliette Lochy and Isabelle Peretz. 2008. Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception* 25: 357-368.
Peretz, Isabelle. 1993. Auditory atonalia for melodies. *Cognitive Neuropsychology* 10: 21-56.
Peretz, Isabelle and Max Coltheart. 2003. Modularity of music processing. *Nature Reviews Neuroscience* 6: 688-691.
Pulvermüller, Friedemann and Luciano Fadiga. 2010. Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience* 11: 351-360.
de Renzi, Enio and Luigi Vignolo. 1962. The Token Test: a sensitive test to detect receptive disturbances in aphasics. *Brain* 85: 665–678.
Rizzolatti, Giacomo and Michael Arbib. 1998. Language within our grasp. *Trends in Neurosciences* 21: 188-194.
Rooth, Mats. 1985. *Association with Focus*, Ph.D. Dissertation, University of Massachusetts, Amherst.
Rooth, Mats. 1992. A Theory of Focus Interpretation. *Natural Language Semantics* 1: 75-116.
Schuell, Hildred. 1965. *Minnesota Test for Differential Diagnosis of Aphasia.* Minneapolis, MN: University of Minnesota Press.
Schnupp, Jan, Israel Nelken and Andrew J. King. 2011. *Auditory Neuroscience - Making Sense of Sound*. Cambridge, MA: MIT Press.
Stewart, Lauren. 2008. Fractionating the musical mind: insights from congenital amusia. *Current Opinion in Neurobiology* 18: 127–30.
Varley, Rosemary A., Nicolai J. C. Klessinger, Charles A. J. Romanowski, and Michael Siegal. 2005. Agrammatic but numerate. *Proceedings of the National Academy of Sciences of the United States of America* 102: 3519-3524.
Venezia, Jonathan and Greg Hickok. 2009. Mirror neurons the motor system and language: From the motor theory to embodied cognition and beyond. *Language and Linguistics Compass* 3: 1-14.

---

1 An earlier version of this paper was presented at *Music and Brains: The Surprising Link*, an ELSC/ICNC conference at Mishkenot Sha'ananim – Jerusalem, the Hebrew University of Jerusalem, February 10th, 2013. I would like to thank the organizers, Eli Nelken, Ronny Granot and Nori Jacoby, for their kind

invitation. I also thank the following agencies and institutions for their support: Edmond and Lily Safra Center for Brain Sciences, Canada Research Chairs (CRC), and the Canadian Social Science and Humanities Research Council (SSHRC). Eli Nelken's comments, and the crucial help of Michael Wagner, Director of McGill's Prosodylab, is also gratefully acknowledged.

2 This evidence, however, is mostly based on work at the single word level, while the linguistic perspective focuses on operations that form larger expressions from more basic units (Varley, Klessinger, Romanowski and Siegal 2005 being a possible exception). See Heim et al. 2012; Deschamps et al. submitted, for further evidence that bears on this issue.

3 Bernstein also proposed to use linguistic tools in order to "build an analogy between musical and linguistic procedures," and to seek the "world-wide inborn musical grammar" (lecture 1).

4 J.C. has courageously participated in a public concert subsequent to her stroke. Her singing in this event can be viewed at http://www.drunkenboat.com/db7/feature-aphasia/curtis/index.html.

5 Note that localizing claims here must be taken as rough approximations, rather than precise pointers, as these authors use functional, as opposed to anatomical, localization (fROIs). Indeed, their expressed focus is on *DD*s of function, rather than on the identification of the exact anatomical loci of these functions.

6 The pitch difference in the language task (5-11 semi-tones) is greater than the one in its musical counterpart, in which, recall, one note was change by a single semi-tone. Yet this difference between the musical and linguistic discrimination tasks cannot be the reason for success, as the same difference did not help the amusical subjects when asked to discriminate question/statement pairs.

7 Patel et al. (2008) and Stewart (2008) mention the well-known economist and public figure Milton Friedman, as well as activist Ché Guevara, as having been amusical. One can't help but doubt the possibility that their deficit extended to linguistic pitch in a manner that that would have hampered their ability to distinguish questions from statements.

8 Notice that the set A is constructed so as to exclude alternatives that *p* entails (Fox 2007). E.g., *p=that John introduced Bill to Sue* entails the alternative *that John introduced someone to Sue*. As the latter is weaker than *p*, it carries the same truth value as *p*. Thus, of the set of possible alternatives, we only include the set of non-weaker (NW) ones, that contains those alternatives to *p* which are not entailed by *p*:

(i) NW(p, A) = $\{q \in$ A: p does not entail $q\}$

E.g., *that John introduced Betty to Sue*, an alternative to *p*, neither entails nor is entailed by *p*. It is thus a member of NW. For notational simplicity, I henceforth assume that Ac=NW. This assumption will become more significant below, in the context of *only*.

9 Once again, just non-weaker alternatives are negated by *only* (Fox 2007) – all others are entailed by *p*, hence true (as *p* is presupposed to be true).

10 A reviewer notes that pitch accent on *Bill* in (19) ensures that *Bill* is focus-marked, but pitch accent on *Sue* in (18) is compatible with F-marking either on *Sue*, or on the whole VP *introduced Bill to Sue* . As empirical evidence, s/he notes that (19) is a good answer to *what did John do yesterday?*, but (18) is not. This observation, while valid, has no interaction with the present proposal: (18)-(19) uncontroversially differ in truth conditions due to focus marking, and the verification task at issue, in which thescenarios mentioned in the text are provided, would therefore distinguish between the 2 approaches to amusia as described in the text.

11 http://www.youtube.com/watch?v=RWsx1X8PV_A