Conflicting constraints in language and music

PhD-candidate:	**Maartje Schreuder**
Masters degree	**General Linguistics, RUG Groningen**
Project name	**Prosodic processes in language and music**
Project period	**September 2000 - September 2004**
Supervisors	**Prof. Dr. Jan Koster (promoter) (RuG, Dept. of Linguistics)** **Dr. Dicky Gilbers (RuG, Dept. of Linguistics)**
Address	**University of Groningen** **Department of Linguistics** **Oude Kijk in 't Jatstraat 26** **P.O. Box 716** **9700 AS Groningen** **e-mail: M.J.Schreuder@let.rug.nl** **phone: 050-3635920**
Reviewers	**1 Prof. Dr. Theo Kuipers** **2 Dr. Fred Keijzer** **3 Drs. Leonoor van der Beek** **4 Ir. Gerke Hoiting**

Contents

Abstract

The aim of this paper is to embed my project in the more general research programs of the BCN-disciplines, in the framework of philosophy of science and mind. With this paper I intend to introduce the background theories of my project, which are Generative Linguistics (Chomsky, 1965), Optimality Theory (Prince & Smolensky, 1993) and a Generative Theory of Tonal Music (Lerdahl & Jackendoff, 1983). I will make clear what the similarities are between Optimality Theory and the Generative Music Theory, and between the subjects of the two theories, language and music. The aim of my project is to show that language and music are very similar in structure, and to investigate whether music theory can give some explanations for linguistic issues. I will give an example of a linguistic issue, concerned with rhythmic structure, and show how music can help out in this case. I will conclude with some *Gestalt* psychological grouping principles, which are originally derived from the visual system, but I will show that those principles also do apply in language and in music, and probably in many other temporally ordered systems, like dance (Lasher, 1978), for example.

1. Introduction

In linguistics there is one dominant research program, which is Generative Grammar (Chomsky, 1965). In this program, language is seen as an autonomous, cognitive capacity. In this program 'grammar' means the description of the (mostly unconscious) knowledge of a speaker of a language about that language. Cognition is assumed to be for the most part unconscious.

Linguists usually divide the language faculty into several modules, which are semantics, pragmatics, syntax, lexicology, morphology and phonology. In processing or producing language all those modules work together to get the whole linguistic picture of an utterance or text. Phonology is the module that studies the (abstract) structure of speech sounds, which includes stress (accent), rhythm and intonation. For the study of sounds, phonologists work together with phoneticians, who are the physicists in linguistics. Phoneticians do not study the linguistic cognition, but mere linguistic behaviour and physiology. Generative linguists are interested in cognition, but lately a new co-operation has emerged between phonologists and phoneticians (Laboratory Phonology), because the *parole* is indispensable in studying the *langue*.

I am a phonologist, and my project is mainly concerned with rhythm and (intonational) phrasing. But I have also some experience in the musical world, and I am interested in the link with musical theory. Although there seem to be many similarities in the syntactic module, the similarities with the phonological field are more obvious, as rhythm, phrasing and melody are also properties in music theory. I use phonetic devices to analyse tape-recorded utterances (and musical pieces). From those results I will distill phonological conclusions.

2. Description of the Research Programs

2.1 Generative Linguistics

Generative Linguistics is a theoretical program which states that humans are born with a so-called Language Acquisition Device. The basic idea of Generative Linguistics is that all languages are based on a Universal Grammar (UG). This UG is a scheme that delimits the learnable language systems (grammars), with some specific choice options. Every single language is just a variant of that universal grammar. Language is thus seen as an autonomous cognitive capacity, for which our brains are specialized. It is 'generative' because it is a system of explicit rules which derive output structures from the basic input structures. (1) shows an example of what such rules do. This is an example from Turkish, a language in which case is derived by adding morphemes to a stem.

(1) Alternations in Turkish (Clements & Keyser, 1983; cf. Burzio, 1995)

	Accusative	Nominative	Ablative
Degemination:			
`feeling'	hiss+i	his	his+ten
`right'	hakk+i	hak	hak+tan
Epenthesis:			
`transfer'	devr+i	devir	devir+den
`abdomen'	karn+i	karin	karn+dan
Vowel shortening:			
`time'	zama:n+i	zaman	zaman+dan
`proof'	isapa:t+i	ispat	ispat+tan

This example shows that the stem (the Accusative form in front of the + sign), i.e. the input, can undergo three kinds of rules to come to the other case forms, i.e. the output forms: degemination, epenthesis and vowel shortening. From the stem ‘hiss’ the geminate consonant is degeminated to ‘his’, in the example ‘devr+i’ the ‘r’ from the stem exchanges with the *i* sound in the nominative and ablative forms, and from the stem ‘zama:n’ the long ‘a:’ sound is shortened in the alternations. This is a nice description of what happens, but it does not explain why these three different processes occur. For reasons of this kind, a new theory emerged.

2.2 Optimality Theory

Since 1993 there is a new linguistic theory, based on generative linguistics and connectionism: Optimality Theory (OT) (Prince & Smolensky, 1993). This theory, first introduced in phonology, is now one of the dominant theories in phonology and it is expanding more and more from phonology to other linguistic disciplines. OT works with well-formedness constraints on outputs. So this theory does not look at derivational rules from input to output, but is only interested in the output-forms. According to this theory, all the processes have a certain goal, which is to get an optimal output form. When we look at example (1) above, we observe that the output forms all have a CVC (C = Consonant, V = Vowel) syllable form, which is (together with CVV) the preferred form in Turkish. Now the different rules are not important any more, what counts is the result of those rules. Some outputs are more preferred than others. Note that this is still generative, because the theory still starts with inputs, but the form of the inputs is irrelevant. This is a totally new point of departure and it is also the strong point of the theory. Which kinds of structures are preferred is stated in universally formulated constraints. Well-formedness constraints determine grammaticality. These unordered, universally formulated constraints together form UG. At least an important subset of all constraints is shared by all languages, forming part of Universal Grammar.

Also new is that these constraints are not strict claims on outputs. They are potentially conflicting and violable. It is even possible for an optimal output form to violate a certain constraint. This is only possible, however, if violation of that constraint leads to the satisfaction of a more important constraint. Individual languages rank the universal constraints differently in language specific hierarchies in such a way that higher ranked constraints have total dominance over lower ranked ones. The examples in (2) show two OT tables of stress application, in which two constraints are competing. The output forms depend on the place of the relevant constraints in the hierarchy of the language in question. The output form which does not violate the higher-ranked constraint wins, even if the lower constraint is violated by that form.

(2) Conflicting constraints:

Nonfinality: stress never falls on the last syllable
Peak prominence: stress falls on the heaviest syllable (in this case: /a/ is longer than /ɑ/, and therefore heavier)

a.

constraints → ↓output candidates	nonfinality	peak prominence
´☛ pá.pa		*
pɑ.pá	*!	

b.

constraints → ↓output candidates	peak prominence	nonfinality
´ pá.pa	*!	
☛ pɑ.pá		*

* this constraint is violated by the output candidate
! this violation is fatal.
▭ the choice has been made, no further analysis necessary
☛ this is the optimal candidate

In the hypothetical language in (2a), of which Dutch could be an example, stress is preferred not to fall on the last syllable. The second output candidate has its stress on the last syllable and it therefore violates the constraint ‘nonfinality’. The first candidate violates the constraint ‘peak prominence’. However, nonfinality is higher ranked in this language, so the form ‘pɑ́pa’ is better than ‘pɑ.pá’. In language (2b), it is just the other way around.
The idea of conflicting constraints was new in linguistics, but Optimality Theory had a predecessor in another area of science: music theory, and it owes a great deal to that work, which is the work of Lerdahl and Jackendoff (1980, 1977, 1983).

2.3 A Generative Theory of Tonal Music

In their generative theory for tonal music Lerdahl and Jackendoff (1983) describe how a listener (mostly unconsciously) constructs connections in the perceived sounds. The listener is capable of recognizing the construction of a piece of music by considering some notes / chords as more prominent than others. This enables him for example to compare various variations on one theme and to relate them to the original theme. It enables him to get to the bottom of the construction of a complete piece, as well as the constructions of the different parts of that piece.

Jackendoff and Lerdahl (1980) point out the resemblance between the ways both linguists and musicologists structure their research objects. This insight gave rise to the proposal of a formal generative theory of tonal music (Lerdahl and Jackendoff 1983), in which they describe musical intuition. Above all, insights from non-linear phonology (*cf.* Liberman 1975; Liberman and Prince 1977; Selkirk 1984; Hayes 1984 a.o.) led to scores provided with tree structures, indicating heads and dependent constituents in the investigated domains, which is the usual way of representing dependency relations in linguistics. In this way, composer Lerdahl and linguist Jackendoff bring to life a synthesis of linguistic methodology and the insights of music theory. In both disciplines the research object is structured hierarchically and in each domain the important and less important constituents are defined. In Lerdahl and Jackendoff's music theory, these heads and dependents are defined by preference rules determining which outputs, i.e., the possible interpretations of a musical piece, are well-formed. Some outputs are more preferred than others. Preference rules, however, are not strict claims on outputs. This makes the preference rules very similar to the OT-constraints. (3) gives three conflicting musical preference rules for the time span reduction.

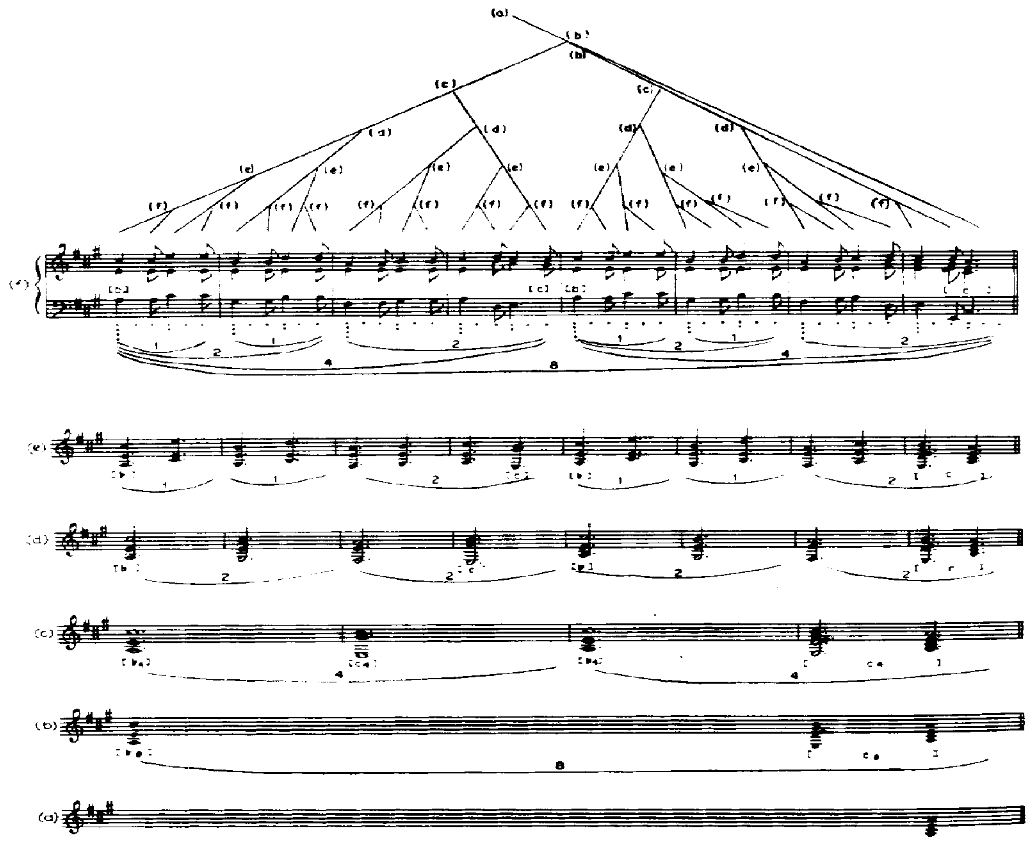
(3) Conflicting preference rules

TSRPR 1: Choose as the head of a time span the chord (or the note) which is in a relative strong metrical position (= the 1st position in a measure)

TSRPR 2: Choose as the head of a time span the chord (or the note) which is relatively harmonically consonant.

TSRPR 7: Choose as the head of the time span the chord (or the note) which emphasizes the end of a group as a cadence .

(4) Time span reduction (Lerdahl & Jackendoff, 1977)



In the time span reduction in (4) the musical surface is analysed to the bare basis. Of each group at all levels the most important chord is chosen by means of the time span reduction preference rules (TSRPR's). For this part we can determine the heads, by means of application of the TSRPR-hierarchy. The first four measures of the piece form the first group. In measure 4 the A^6-chord is the most stable (harmonically consonant) chord, and thus the head. In measure 4 the E-chord is the head, because it marks the end of the whole group of four measures. Now the head has to be chosen for the group which is formed by measures 3 and 4 together. Metrically speaking, the A^6-chord is still the strongest. But in tonal music TSRPR 7 dominates TSRPR 1. In this example we can recognize the same kind of constraints as in the linguistic stress application example, in which a positional constraint conflicts with a constraint for which the quality of a segment counts. In (5) we give an example of an OT-like musical analysis. Although the A^6-chord is metrically speaking in a stronger position than the E, the dominant RSRPR 7 prefers the dominant chord E as the cadence in this phrase.

(5) OT analysis (Gilbers & Schreuder, to appear)

constraints → A^6 – E Candidates ↓	TSRPR 7	TSRPR 2	TSRPR 1
☞ E			*
A^6	*!	*	

With this very brief introduction in the main research programs of my project, I hope to have shown that the basic structures of language and music are quite similar and that they can be analysed in comparable ways.

3. An interdisciplinary project: language and music

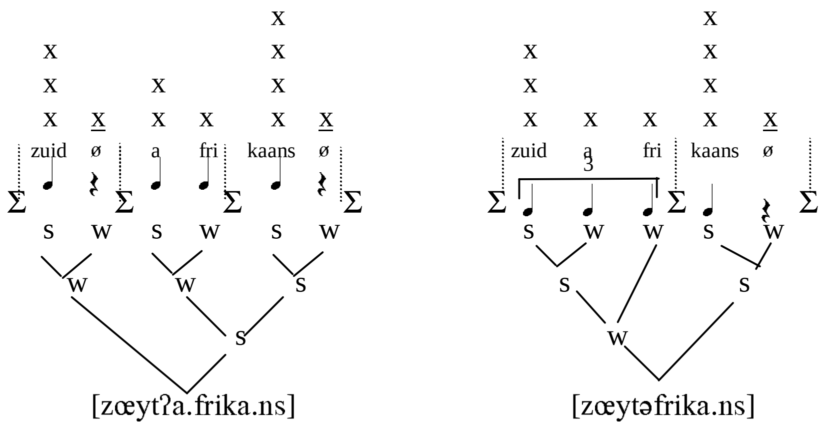
This research project is an interdisciplinary project. It is linguistically based, but I investigate whether cross-pollination between linguistics and musicology is possible. Generative linguistics is a theoretical program which states that humans are born with a universal grammar. Every single language is just a variant of that universal grammar. Language is thus seen as a cognitive capacity, for which our brains are specialized. I have shown that the musical theory of Lerdahl and Jackendoff is very similar to the Generative theory OT. Music is also a universal cognitive capacity and language and music share many features. First of all, both language and music are 'made of' sound, and moreover, these sounds are structured in a hierarchical way. They consist of entities like phonemes or notes, words or motifs, and sentences or musical phrases of which a text or piece is built hierarchically. Those linguistic and musical entities have much in common, for example, they share rhythm and pitch, forming linguistic intonation patterns or melodies. Because of the many structural similarities, it is interesting whether there are linguistic processes which can be explained by music theory. One process I am interested in, is rhythmic variability.

3.1 Rhythmic variability

In fast speech variable rhythmic patterns can occur for the same phrase. An important issue in my phonological research is the question whether the influence of a higher speech rate leads to adjustment of the phonological structure, or just to 'phonetic compression'. Phonetic compression is mainly shortening and merging of vowels and consonants, with preservation of the phonological (rythmical and metrical) structure. If the phonological structure is adjusted, this means that every speech rate has its own register. So in a fast register it would be important to keep the rhythm as simple and regular as possible, in order for nothing to hinder fast speaking. To give an idea of the relevant data, I give an example of rhythmic variability in the Dutch word *zuidafrikaans* ‘South African’. This example illustrates how music theory can solve a linguistic issue, and that the music theory is a ‘supplying’ research program to the linguistic ‘guide’ research program. When we look at the rhythm of the word *zuidafrikaans*, we see that it consists of four syllables and at a normal speaking rate (andante), it has three beats, at the syllables *zuid-*, *-a-*, and *-kaans*. These syllables are in strong metrical positions, i.e. each is the first syllable in a foot (Σ). In the fast variant (allegro) the number of strong beats or the number of feet can be changed from three into two. Evidence for this restructuring can be found by looking at the place of reduction possibilities in the phrase. Generally reduction of a vowel to schwa is only possible in weak syllables. In the s s w s structure (s = strong, w = weak) of *zuidafrikaans* in a lento tempo, *-a-* cannot be reduced. In fast speech this is possible, however. This indicates that restructuring occurs, by which the position of that syllable becomes weak. That is possible if the rhythm is simplified to a triplet, in which only the first note is strong. In the weak position reduction of the syllable *–a-* to schwa is possible. Like in music a faster tempo apparently does not always lead to ‘phonetic compression’, but often to restructuring of the rhythm. This process is illustrated in example (5).

(5) Restructuring *zuidafrikaans* (arboreal, musical and in grid form) (after Gilbers, 1987)

a. adagio b. allegro

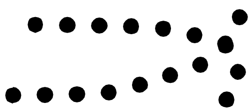


This example is just one of the interesting linguistic issues I hope to explain with music theory.
In the next section I will show that the similarities between language and music are for a great part reducible to grouping principles, and that those grouping principles are not just linguistically or musical, but that they do also apply in the visual field, for example.

3.2 General grouping principles: cognition

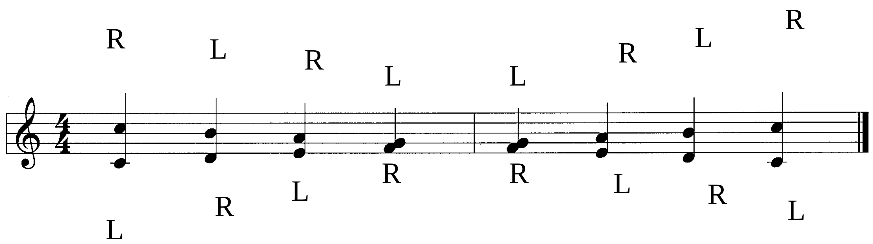
In Gestalt Psychology grouping principles are formulated according to which we structure the visual world. And similar to the OT-constraints and the musical preference rules, these principles can be in conflict with each other. The grouping principles are formulated in (6):

(6) Visual grouping principles from Gestalt Psychology (Deutsch, 1982)

Proximity:	elements that are close to each other, are perceived as a group;	●● ●● ●● ●●
Similarity:	elements that are similar are perceived as a group;	* ● * ● * ● * ● * ● * ●
Good Continuation:	elements in a sequence following each other in a certain direction are perceived as a group;	
Common Fate:	elements that move in the same direction are perceived as a group.	

Deutsch (1982) demonstrates that these principles apply also to music, and I will briefly illustrate that it is a linguistic way of grouping as well. Deutsch describes musical experiments in which the listeners are wearing earphones. Different notes are presented to the left and the right ear. I will show some examples to illustrate some of the principles and how conflicts are solved. In (7a) two scales are played simultaneously, one from a high c to the low c and the other from low c to high c. The notes of the scales switch back and forth from one ear to the other.

(7) a. stimulus

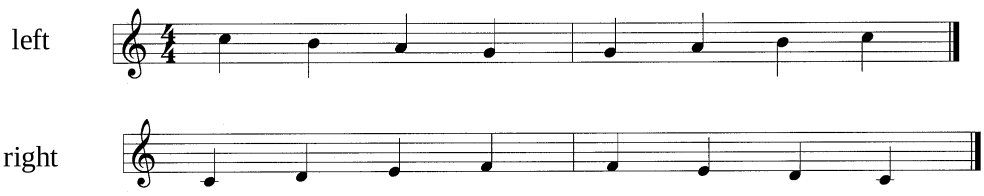


The subjects were asked what they perceived. A possible perception could be the two scales, one scale to the left ear and the other to the right ear, as illustrated in (7b). If this was the case, the principle of Good Continuation would be the winner.

(7) b. possible percept (Good Continuation)

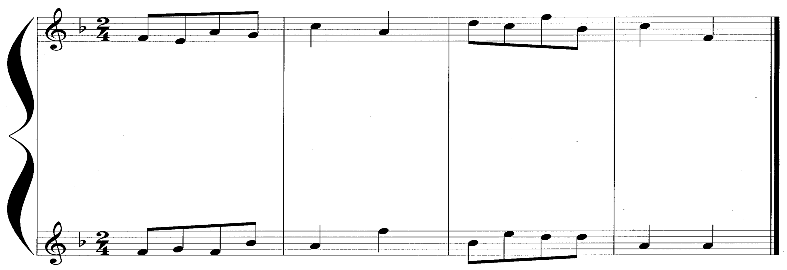
But this was not the case. Instead, it was perceived as in (7c). The notes were grouped on the basis of frequency range, i.e. the high tones formed a group and the low tones formed the other group. This means that the principle of Proximity wins. And even under normal, stereo listening conditions, without the switching earphones, it is very difficult to perceive it as two scales.

(7) c. actual percept (Proximity)

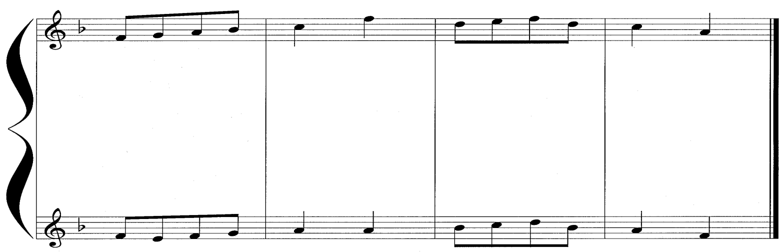


In (8) another example of Deutsch is given. It shows two melodies (8a), one presented to the left ear and the other to the right. But the perception is as in (8b), and it is almost impossible to perceive it as in (8a). This is an illustration of the principle of Good Continuation, because in (8b) the sequences of notes form nice upward and downward melodic lines.

(8) a. stimulus

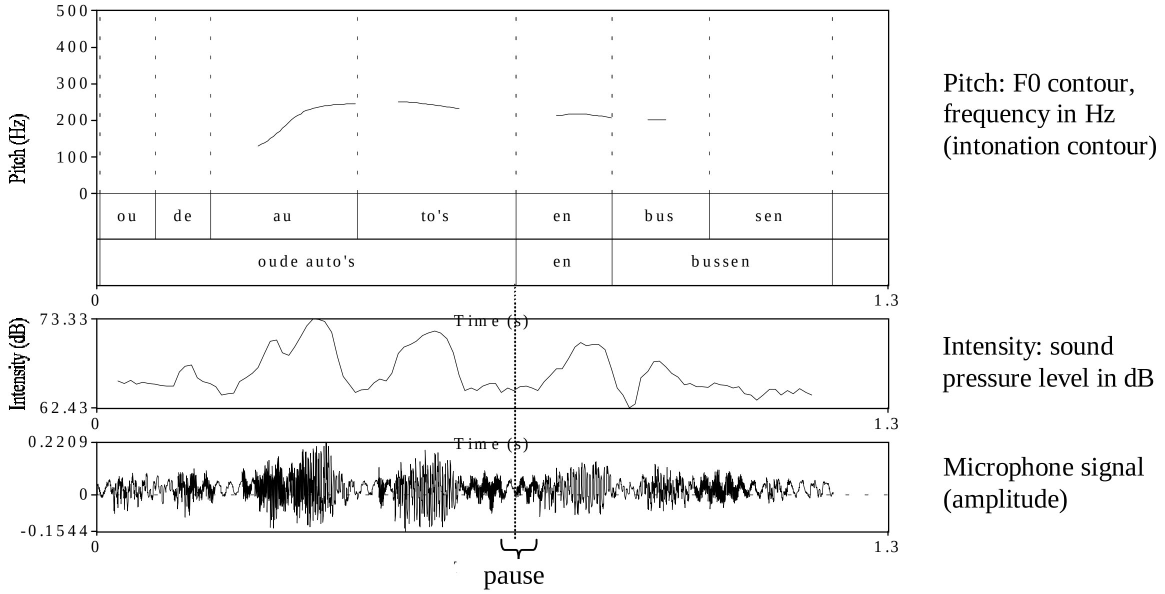


b. percept: Good Continuation

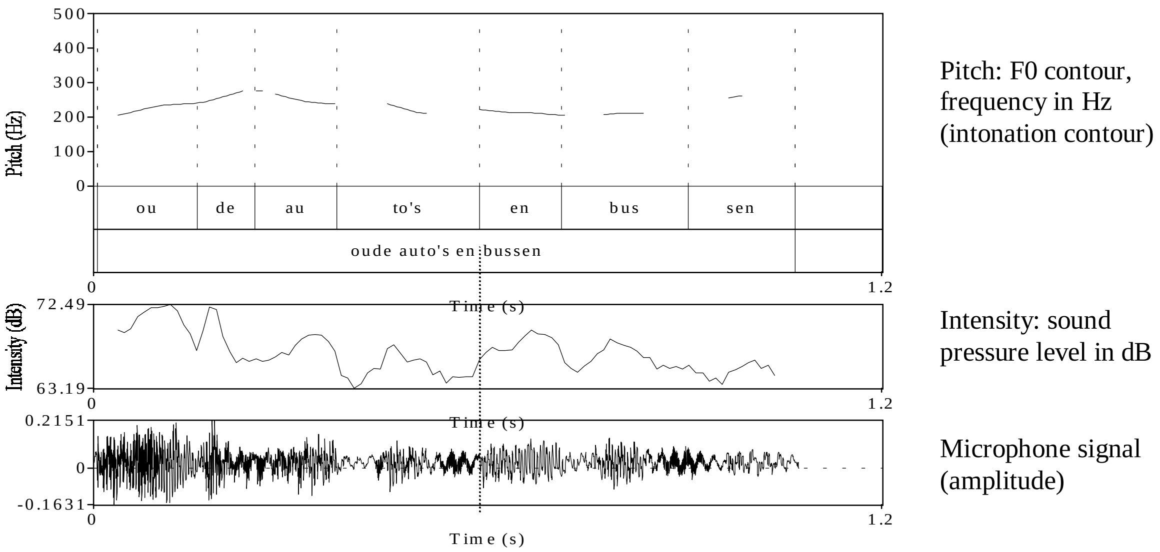


In language similar grouping principles can be observed. Proximity, for example, can be seen in the fact that phrases are delimited by pauses. All the words that are close to each other, i.e. without pauses between them or bound by all kinds of linking processes, form a group. Example (9) illustrates the ambiguous phrase *oude auto's en bussen* 'old cars and buses', which can be interpreted as either *there are old cars and there are buses* or *there are old cars and old buses*.

(9) a. Disambiguation: grouping principle of Proximity

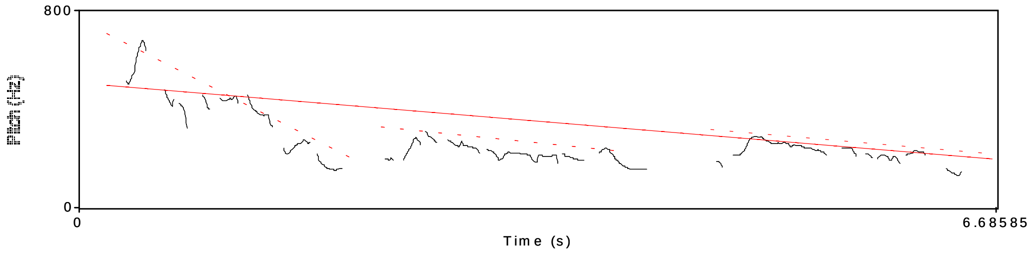


b.



From example (9a) we can read a pause in the intensity and the amplitude (these signals contain a large amount of noise, therefore silence is not zero in these pictures) between the words *auto's* and *en bussen*, in the area around the dotted line. In (9b) there is hardly any pause (the energy-less span before *en* in (9b) is due to the low energy level of the s sound, as it is to the low energy span before the pause in (9a)) and the intonation contours are also quite different. In (9b) the contour forms a gradually declining slope over the whole phrase, indicating that this is a single phrase, whereas the contour in (9a) shows a less gradual course. This declination over phrases is an illustration of a second grouping principle: Good Continuation (or Common Fate). This can be illustrated even better with the F0 contours in the example in (10), which shows that a sentence is a group with a general declining tone, but that the subsentences form subgroups with their own slope.

(10) Declination slopes, Good Continuation / Common Fate



Joost gaat op vakantie naar Frankrijk | en Anton gaat met de zeilboot naar Engeland | en Karel wil naar een cursus in Duitsland.

(Schreuder, 1999)

These examples of grouping principles in the visual field, in language and in music show that the grouping principles are very general in nature. When we look at processes like these, we ask ourselves why these processes occur, and why they occur in two instances of our cognitive system. My assumption is that our cognition structures all temporal processes in a similar way, which is by grouping and contrasting elements in a certain manner. This idea partly departs from the central idea of generative linguistics that we have a language specific learning system and that language is autonomous, because it means that we have general structuring capacities applying to language as well as to music and other capacities.

4. The research programs in philosophical terms

The program is partly descriptive and partly explanatory. The linguistic part is largely descriptive. The musical part can provide the explanations for the phonological issues. The linguistic and the musical conclusions together are meant to give explanatory evidence for general cognitive capacities concerning temporally ordered behaviour.

One could also say that my view is a reductionist view, in that it reduces language and music to cognition. I would say that it is useful to regard language and music, as well as other human cognitive capacities, as separate but co-operating instances of our cognitive system, but it is still important to look for specificities of the separate instances. This means that both the top-down and the bottom-up ways of investigating human cognitive capacities are relevant. To call it 'reduction' is only a matter of terminology, and therefore it seems irrelevant to me.

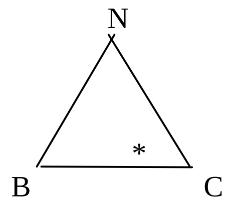
5. Interactions and co-operation

This project is interesting from many points of view, like psychology, because of the general cognitive aspects. It is interesting from a neurological point of view because linguistic and musical processing share many cerebral areas (Borchgrevink, 1982; Maess, et al., 2001), and another possible cooperation would be computational linguistics and computational musicology, using linguistic-like grammar rules to generate, describe or predict musical phrases. Many of these areas of research could give a fruitful contribution to this project. In fact, I planned to do some experiments together with colleagues working on brain imaging techniques (Stowe, Haverkort) and another plan is indeed to combine computational linguistics and computational musicology, using syntactic rules (with Menno van Zaanen).

6. Localization in the BCN triangle

Linguistics studies are interested in linguistic cognition, and in order to get information about the cognitive linguistic skills and processes, linguistic behaviour is studied because cognition itself is largely invisible. This means that my project is mainly cognitive in nature, but there is a behavioural aspect, which is necessary for the analysis. The place of my project in the BCN triangle is as indicated in (11):

(11)



Bibliography

Borchgrevink, Hans M. (1982). Prosody and Musical Rhythm are Controlled by the Speech Hemisphere, in: Clynes, Manfred (ed.), *Music, Mind and Brain: The Neuropsychology of Music*, Plenum Press, New York.

Burzio, L. (1995). The Rise of Optimality Theory. Glot International, 1 (6): 3-7.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.

Clements, G.N. & Keyser, S.J. (1983). *C-V Phonology*. Cambridge, Mass: MIT Press.

Deutsch, Diana (ed.) (1982). *The psychology of music*. Academic Press series in cognition and perception, New York: Academic Press.

Gilbers, D. (1987). Ritmische Structuur. In: *Glot* 10: 271-292.

Gilbers, D. & M. Schreuder (to appear) Language and music in Optimality Theory. Ms., University of Groningen.

Hayes, B. (1984). The Phonology of Rhythm in English. In: *LI* 15. 1: 33-74.

Lasher, M. (1978). A Study in the Cognitive Representation of Human Motion. Ph.D. dissertation, Columbia University.

Lerdahl, F. & R. Jackendoff (1977). Toward a Formal Theory of Tonal Music, *Journal of music theory*, vol. 21: 111-171.

Lerdahl, F. & R. Jackendoff (1981). On the Theory of Grouping and Meter, *The musical quarterly*, vol. 67: 479-506.

Lerdahl, F. & R. Jackendoff (1983). *A Generative Theory of Tonal Music*. The MIT Press, Cambridge, Massachusetts, London, England.

Liberman, M. (1975). *The Intonational System of English*. Garland Publishing, Inc., New York & London.

Liberman, M. & A. Prince (1977). On Stress and Linguistic Rhythm. *Linguistic Inquiry*, 8 (2): 249-336.

Maess, B.; Koelsch, S.; Gunter, T.C. & A.D. Friederici (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, 4 (5): 540-545.

Prince, A. & P. Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms., Rutgers University Centre for Cognitive Science, Technical Report 2.

Schreuder, M. (1999). *Taal en Muziek, de structurele tweeling van ons cognitieve systeem*. Doctoraalscriptie, Groningen.

Selkirk, E.O. (1984). *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge, Mass.: MIT Press.