Syllables and clauses in language – notes and phrases in music

Gertraud Fenk-Oczlon & August Fenk
Alps-Adriatic University of Klagenfurt

Conference on "Language and Music as Cognitive Systems", University of Cambridge, 13-17 May 2007

Overview

- **We start with the method and the results of our crosslinguistic studies regarding the "size" and the rhythmical organization of clauses**
- **and relate the results to data concerning the "size" and the rhythmical organization of musical phrases.**
- **A further parallel regards the inventory size: vowels in language – notes in music**

Overview

The parallels described will be interpreted from an evolutionary perspective emphasizing

- **"half-musical unanalysed utterances"(Jespersen 1895) as the origin of both language and music**
- **the role of determinants of intonation units shaping clauses in language, phrases in vocal music, and – indirectly – also in instrumental music.**

Information processing limitations in LANGUAGE

- Hypothesis:

  The number of syllables per clause will vary within the range of the magical number seven.
- Method:

  Native speakers of 33 (17 Indo-European, 16 non-Indo-European) languages translated 22 German clauses into their mother tongue and determined the length of their translations in syllables and in words.

Sentences

(1) The child is waiting for its meal.	**(12) My girlfriend is helping me.**
(2) The sun is shining.	**(13) My brother is a hunter.**
(3) Blood is red.	**(14) The water is cold.**
(4) I think of you.	**(15) The dog is outside.**
(5) Our neighbour is a farmer.	**(16) My father is a fisherman.**
(6) She trusts her friend.	**(17) Grandfather is sleeping.**
(7) She sings.	**(18) A mother loves her son .**
(8) A father looks after his family.	**(19) Aunty is at home.**
(9) The girl is industrious.	**(20) My sister is collecting wood.**
(10) I thank the teacher.	**(21) He is building a hut.**
(11) The spring is on the right.	**(22) It´s raining.**

Results

- The mean number of **syllables** per clause was found to be located almost exactly in the range of Miller's 7 plus minus 2 elements:

  The lowest size was a mean of 5.05 syllables (Dutch), and only Japanese with 10.2 syllables per clause was located outside the hypothesized range of 5-9 syllables
- The mean number of **words** was about 4, ranging from 2.5 in Arabic to 5.4 in Chinese (→ Cowan's 4 plus minus 1)

The mean number of syllables per clause in 34 languages

Dutch	5.05										
French	5.32										
Czech	5.36										
Chin.	5.41										
Hebr.	5.46										
Slov.	5.50										
Germ.	5.50	Bamb.	6.46								
Icel.	5.50	Turk.	6.46								
Eston.	5.68	Alban.	6.55								
Russ.	5.68	Yoruba	6.59								
Croat.	5.77	Port.	6.64	Hopi	7.12						
Engl.	5.77	Pers.	6.64	Nava.	7.41						
Ewon.	5.77	Hindi.	6.77	Ital.	7.50	Korean	8.18				
Hung.	5.91	Pan.	6.77	Greek	7.55	Anjang	8.23				
Arab.	5.96	Mac.	6.96	Span.	7.96	Basque	8.27	Chiqu.	9.14	Japan.	10.23

Syllable complexity as decisive factor

- Dutch is known for its complex syllables (e.g. CCVCC).
- Japanese is known for its simple syllable structures (CV or V).
- A statistically significant negative correlation between number of phonemes per syllable and number of syllables per clause.

  $r = -0.77$, $p < 1\%$

Tempo
in stress-timed (1-4), syllable-timed (5-6), and mora-timed (7) languages

	a syll/clause	**b** phon/syll	**c** vocalic nPVI
(1) Dutch	5.05	2.97	65.5
(2) German	5.50	2.84	59.7
(3) English	5.77	2.69	57.2
(4) Thai	5.29	2.51	65.8
(5) French	5.32	2.47	43.5
(6) Spanish	7.96	2.09	29.7
(7) Japanese	10.23	1.88	40.9

Duration of Syllables

- Kegel (1990) concludes from a series of psycholinguistic experiments that 100 to 500 msec are necessary for the processing of syllables
- Chu& Feng (2001) analyzed a corpus of 13,000 sentences of Mandarin. They report 245 msec as the overall mean duration of syllables; more than 99.5 % of all syllables were between 100 and 500 msec

Size of clauses

- The mean duration of clauses is about 2 sec. Chafe (1987): "new intonation units typically beginn about two seconds apart."
- 2 seconds correspond to 10 short syllables or 5 complex syllables

  2 seconds ↔ 5 – 10 syllables

Information processing limitations

shaping the structure of language

- **The temporal span perspective:**
  ~1.5 – 3 sec seem to be a temporal span for processing information
  → psychological present (Fraisse 1957, Pöppel 1986)
  → Baddeley's phonological loop: ~ 2sec
  → an increase of the Weber fraction between ca 1.5 – 2 sec (Lavois & Grondin 2004)
- **The perspective of limited capacities:**
  7 plus minus 2 (Miller 1956)
  4 plus minus 1 (Cowan 2001)

Time spans in MUSIC

- **A time span of ~200 msec to ~1,8 sec is critical for the processing of music (e.g. Fraisse 1957, Justus & Bharucha 2002)**
  - within this time-span we tend to group an isochronous pattern in twos, threes or fours (Meumann 1894)
  - within this time-span an accurate synchronization to an isochronous pulse works (Fraisse et al.1949)
  - beyond ~ 1.8 seconds subjective rhythmization becomes impossible: The tick and the tuck are no longer perceptually linked; melody disappears and is substituted by single notes

Time spans and capacity limits in music

- The duration of musical phrases roughly corresponds to Fraisse's psychological present. (Parncutt & Pascall 2002)
- 30 to 300 pulses per minute (Fraisse 1982).
  This amounts to a maximum of **~10 pulses within 2 seconds**

Phrase length in music

The results of a statistical analysis (Huron 1996) of the ***Essen Folk Song Collection***:

- 8 notes as the median phrase length
- 50 % of the phrases had 7 – 9 notes
- 75 % of the phrases had 6 – 10 notes

Phrase length in music

The results of a statistical analysis (Temperley 2001) of the ***Ottmann Collection*:**

- a mean of 7.5 notes per phrase
- 75 % of the phrases had 6 – 10 notes
- less than 1% of the phrases had fewer than 4 or more than 14 notes

Musicians are reading ahead

- The maximum amount of reading ahead varies between 6.8 and 8 notes (Sloboda 1982)
- The note is played about two sec after being read (Sloboda 1982)
  In good sight readers, the eye-hand span (EHS) comprises 2 sec and 7 notes

The more, the faster

- the more single movements per action unit, the shorter the single movement (Schleidt 1992)
- the more syllables per clause, the fewer phonemes per syllable (Fenk-Oczlon & Fenk 1985)
- in phrases containing many notes, the notes are usually very fast (Temperley 2001)

General Results

- The interval of 0.2 to 2 sec is not only relevant in language processing; it also shows in the segmentation of music and in making music
- The maximum number of syllables (per time interval) corresponds with the maximum number of notes: up to a mean of ~ 10 syllables per clause in language and up to a mean of ~ 10 notes per phrase in music
- The more elements per unit (clause, phrase), the shorter the elements

Parallels in the inventory size? Vowels in language - notes in music

- The sound (the sonority) of syllables mainly comes by the vowels constituting the sonority peaks of syllables. Thus the vowels are particularly relevant for vocal music and for music in general. Therefore we expected a coincidence between the number of musical notes and the **number of vowels** - instead of the total number of phonemes (as in Rakowski 1999).

Sheer coincidence?

- **5-vowel** systems are most frequent in languages, and far more languages have **5** or **7 vowels** than have **4** or **6**. (Crothers 1978, Ladefoged 2001)
- pentatonic (**5-note**) scales are used more widely than any other formation. The hepta-tonic (**7-note**) scale is a widespread pattern, too. The hexatonic (**6-note**) scales appear rather rarely in folk music and non literate cultures. (Encyclopaedia Britannica)

Sheer coincidence?

- **12 vowels** (e.g. in Pacoh) are a maximum number according to Crothers. Ladefoged claims a higher number probably because of considering interindividual differences within a language community: e.g. up to 20 vowels in „BBC English“
- The chromatic (**12-note**) scale seems to mark an upper limit. The existence of a higher number of tones (e.g.in the Arab-Persian system) is, at least as a standard, a rather controversial question (Burns 1999)

Discussion: evolutionary perspectives

All the following possibilities might explain the parallels in segmentation and inventory size:

- **A (co-evolutionary) convergence of originally independent systems**
- **Language evolved from music (Darwin 1871)**
- **Music evolved from language**
- **Both evolved from “half-musical unanalysed expressions“ (Jespersen 1895). This idea of the evolution of both language and music from “musilanguage“ (Brown 2000) and/or from “Hmmmmm“ (Mithen 2005) corresponds with descriptions (e.g. Cross 2003) of the “protomusical communication“ between mother and infant**

Discussion

The (less speculative?) core of our explanation:

- **“Singing“ (in a broader sense of the word) is at any rate a very old achievement**
- **Talking as well as singing comes about in intonation units. One has to assume that any determinant of intonation units will be reflected in language as well as in music. Relevant determinants are the “clausal structure” of the breath cycle and the (coordinated) “clausal structure” of those cognitive processes programming the shape of sound.**

Discussion

- **In a very common (and very ancient ?!) form of music, instrumental music accompanies – or is accompanied by – singing, i.e. vocal music. Thus the determinants of intonation units will – indirectly – also shape the rhythmical organization of instrumental music.**

Discussion

We “explained“ the parallels between language and music first of all by rhythmical properties of intonation units. But these units are only a special case of action units (Fenk-Oczlon & Fenk 2002). From this perspective one may assume that the rhythmical commonalities of talking and singing, of vocal and instrumental music, of drumming and dancing – that all these reflect the rhythmical organization of action units and of the (kinestethic) coordination between subunits of our sensori-motor system.

Aus Darwin „The descent of Man – chapter XIX:

As we have every reason to suppose that articulate speech is one of the latest, as it certainly is the highest, of the arts acquired by man, and as the instinctive power of producing musical notes and rhythms is developed low down in the animal series, it would be altogether opposed to the principle of evolution, if we were to admit that man‘s musical capacity has been developed from the tones used in impassioned speech. We must suppose that the rhythms and cadences of oratory are derived from previously musical powers.

...Darwin

..., and as a gibbon, one of the anthropo-morphous apes, pours forth a whole octave of musical notes and may be said to sing, it appears probable that the progenitors of man, either the males or females or both sexes, before acquiring the power of expressing their mutual love in articulate language, endeavoured to charm each other with musical notes and rhythm.