Musical Pitch in Nonsense Syllables: Correlations with the Vowel System and Evolutionary Perspectives

Gertraud Fenk-Oczlon,[*1] August Fenk[#2]

[*]*Department of Linguistics and Computational Linguistics, University of Klagenfurt, Austria*
[#]*Department of Media and Communication Studies, University of Klagenfurt, Austria*
[1]gertraud.fenk@uni-klu.ac.at, [2]august.fenk@uni-klu.ac.at

ABSTRACT

In a first step we present correspondences between the inventories of vowel systems and musical scales across cultures. These correspondences concern an upper limit of roughly 12 elements, a lower limit of 2-3 elements and the 5-vowel systems and pentatonic scales as the most frequent patterns. Since vowels play a decisive role in generating the sound or the sonority of syllables that in turn are basic units of both speech and singing, these parallels are considered as a possible link between language and music. Furthermore we found correspondences between vowel pitch and musical pitch in Alpine yodellers whose text entirely consisted of nonsense syllables: A succession of the vowels [i] and [o] in meaningless syllables was most often accompanied by a succession of high and low pitch in melody. In other Austrian folksongs containing only several passages of nonsense syllables, the present study shows a similar correlation, i.e. a coincidence of the vowel [i] with the highest pitch in melody. These coincidences between the vowel system and musical intervals shall be discussed in regard of a possible evolution of music and speech out of a half-musical precursor.

I. INTRODUCTION

Singing is assumed to be the most original form of music, maybe even older than drumming and than singing accompanied by rhythm instruments. And singing, alone or accompanied by instruments, still is the most widespread form of music. The terms "singing" or "vocal music" not only apply to songs having meaningful texts. According to Nettl (1954, p. 192) they also apply to songs "which have only meaningless syllables." He reports that in many songs of the Arapahos and other American Indians senseless syllables are common as complete or partial song texts. On the next page of this article he already states that "there is a tendency for high tones to be sung on syllables with front vowels /…/ while lower tones are more apt to be sung on back vowels /…/ This may have a physiological basis, since we know it is easier to sing high tones with front vowels and low tones with back vowels".

The idea of a close evolutionary relationship between language and music traces back as far as Charles Darwin who decidedly argued for music as a precursor of articulate speech: "We must suppose that the rhythm and cadences of oratory are derived from previously developed musical powers." (Darwin, 1871, p.12). Jespersen, however, considered two different possibilities: Language as a descendent of music (Jespersen, 1922), as already in Darwin, and both language and music as descendents of "half-musical unanalysed expressions." (Jespersen, 1895). This idea of a common and less specified precursor of language and music seems to get a revival (cf. Brown, 2000, Mithen, 2005, Fitch, 2009). In a recent study we refer, moreover, to Morley's (2003, p. 149) description of a "progressively, increasingly complex proto-language based on tone-dominated affective social utterances" and take Nettl's (2000, p. 469) description of archaic singing as a model for language evolution:

"Simple musical phrases repeated, modified, and repeated again, are an appropriate production system as well as an ideal carrier current system for transmitting symbols, i.e., elements of a coding system associated with particular meanings by virtue of their rule-based use." (Fenk-Oczlon & Fenk, forthcoming, p. 172).

There are of course physical cues that make us perceive a certain phrase as rather musical (sung) or linguistic (spoken), but interestingly the very same "spoken phrase can be made to be heard convincingly as sung /…/ simply by repeating the phrase several times over." (Deutsch, 2008, p. 8). This finding gives rise to several tentative assumptions: Starting from our archaic singing-model one might speculate that the redundancy of a repetitive melodic Gestalt foregrounds minor but significant changes of the articulatory Gestalt. Maybe the long-term differentiation of singing and speech was accompanied by a long-term differentiation of specialized attention mechanisms that may overrule each other or switch depending on slightly differing cues, comparable to perceptual effects described in Gestalt psychology. If so, the speech-relevant mechanism is the one specialized on an utmost efficient extraction and analysis of referential meaning also from less redundant, non-repetitive utterances. The melodic or prosodic features are used in that process, but recede from focal attention - until the phrase is repeated for several times.

Our own studies concern parallels between music and language that may contribute to the contemporary debate about the evolutionary pathways of these achievements in at least two respects:

(a) Parallels in the length of musical and linguistic phrases: Articulatory mechanisms, as well as the cognitive/perceptual mechanisms involved in the programming and on-line control of intonation units, are effective in both speech and singing. Even more generally: Phrases in language (clauses) can be viewed as a special case of action units (Fenk-Oczlon & Fenk, 2002) reflecting their general pattern: a restricted number of elements within a unit of 1.5 to 3 seconds and a negative correlation between the number and the duration of the elements. This argument may of course be extended to phrases in vocal music, in vocal music accompanied by instrumental music, and in purely instrumental music. Some relevant data: In a sample of 34 languages the mean number of syllables per clause was in the range around Miller's (1956) magical number 7, from 5 in Dutch to 10 in Japanese (Fenk-Oczlon & Fenk, 1999), and 75 % of the phrases in the Essen Folksong Collection (Huron, 1996) and the Ottmann Collection

(Temperley, 2001) contain from 6 to 10 notes. A duration of 2-3 sec is not only a characteristics of our musical and linguistic phrases, but is, interestingly, also reported (Sugiura & Masataka, 1995) for the coos of Japanese macaques. Thus, many parallels between language and music can be explained by perceptual and cognitive mechanisms involved in both speech and music.

(b) The second parallel concerns vowels and musical intervals. This will be the topic of the following chapters.

II. CORRESPONDENCES BETWEEN VOWEL SYSTEMS AND MUSICAL SCALES

Since vowels play a decisive role in generating the sound or the sonority of syllables we assumed (Fenk-Oczlon & Fenk, 2005) and indeed found (Fenk-Oczlon & Fenk, forthcoming) some parallels between the inventories of vowel systems and musical scales across cultures. Most languages have 5-vowel systems and all studies available for us claim that the pentatonic (5-tone) scale is more widely used than any other formation. The second most common inventories are, depending on the source, either the 7-vowel systems and heptatonic scales or the 6-vowel system and the hexatonic scales. As to the upper limits of the respective inventories the maximum is around 12. According to Burns, "the present 12-interval Western scale is probably a practical limit. Any division of the octave into intervals smaller than quarter tones is perceptually irrelevant for melodic information." (Burns, 1999: 257). And in language, the 12 vowel-systems again mark a maximum in Crothers' (1978) sample. Concerning the minimum size of vowel or tone inventories, the data vary between 2 or 3 units (Table 1).

It might be promising to look whether languages with a lower number of vowels also tend to have smaller scale sizes. This seems to be a difficult task because of the lack of a database of scale size across musical cultures. But there are some indications in support of this hypothesis:

The pentatonic scale is a generally, and especially in the eastern Asian music, widespread pattern, and the respective eastern Asian languages tend to have 5-vowel systems. Only recently we found relevant data on Arapaho language and music perfectly fitting this hypothesis: According to "The Arapaho Project" of the University of Colorado and the "Encyclopedia-Arapaho music", Arapaho has 4 vowels and 3 basic diphthongs and scales of 4 to 6 tones. There are, in addition, some relevant data concerning Australian Aboriginal music and language which seem to support our hypothesis: It is generally difficult in the songs of Australian Aboriginals to assign pitches to conventional scales; Sachs (1965) describes their melodies as "tumbling strains" or falling melodies. Nevertheless we see some correspondences between the descriptions of their vowel systems and musical intervals: Most of the Aboriginal languages have three vowels, only those from the Northern Territory have five (Butcher & Anderson, 2008). And Lauridsen (1983) reports exactly from those cultures in the Northern Territory the use of a higher number of pitches.

Table 1. Some cross-cultural coincidences between the size of vowel-systems and musical scales.

Vowel systems	Musical scales
• **Most languages have 5 vowels (Crothers 1978); different authors claim either 6 or 7 vowels as the next most frequent inventory**	• **Pentatonic (5-tone) scales are used more widely than any other formation. It is, depending on the source, either followed by the hexatonic (6-tone) scales or by the heptatonic (7-tone) scales**
• **A minimum of 2-3 and a maximum of 12 vowels according to Crothers (1978)**	• **2-3 tones seem to mark a lower limit (Nettl 2000) and the 12-tones of the chromatic scale an upper limit**

III. CORRESPONDENCES BETWEEN VOWEL PITCH AND MELODIC DIRECTION IN NONSENSE SYLLABLES

It has been known since Meyer (1896) that, other things being equal, vowels have intrinsic pitch. High vowels such as [i] or [u] have a higher fundamental frequency (F_0) than low vowels such as [a] or [æ]. Concerning the formants we know that front vowels such as [i] or [y] have a low first formant (F1) and a high second formant (F2) whereas back vowels such as [o] or [u] have a high F1 and a low F2.

An interesting finding relating melodic to vowel pitch is reported in Hughes (2000). He found that in *shoga*, the Japanese mnemonic/iconic system for representing melodies, the vowels are correlated to melodic direction according to their F2 ordering. For instance: a succession of the vowels [i], showing the highest frequency in F2, and [o] with a low F2 was most often accompanied by a succession of high and low pitch in melody.

A. Vowel pitch in nonsense syllables of Alpine yodellers

Having in mind the sound and vowel patterns of some Alpine yodellers and the freedom of yodellers in combining senseless syllables with successions of notes, we assumed that this type of music is composed according to the pattern found in the *shoga* system.

In search for such a correlation we studied (Fenk-Oczlon & Fenk, forthcoming) all of the monophonic yodellers (n = 15) in Pommer's collection from 1893. The whole text of these yodellers is a succession of nonsense syllables (Figure 1).

The results make inferential statistics unnecessary: In 118 out of 121 cases of a syllable containing an [i] preceding a syllable containing an [o] the melody descended; in one case the melody raised, in two cases the pitches stayed at the same. And in the 133 instances of an [o]→[i] succession the melody ascended with only one exception (equal pitch). Instances of an [i] followed or preceded by other vowels were less frequent but offered the same picture: one exception in 44 cases of an [i]→[a] succession, and no exceptions in 10 cases of [a]→[i] and in 6 cases of [u]→[i].

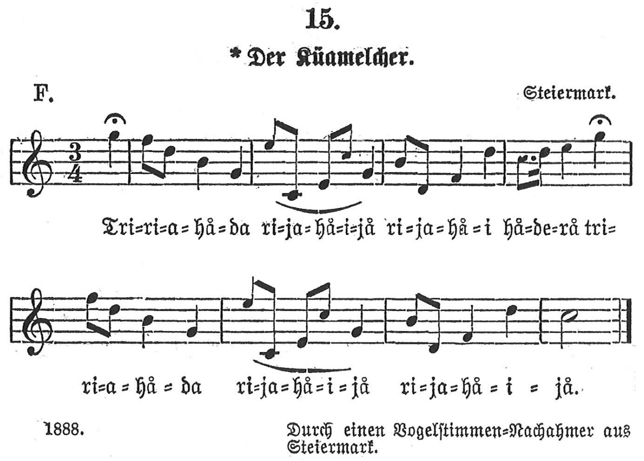

Figure 1. An example of an Alpine yodeller (from Pommer's collection, 1893, p. 9) showing descending melody in [i]→[o,a] successions and ascending melody in [o,a]→[i] successions.

B. Vowel pitch in nonsense syllables of other Austrian folksongs

Can this pattern, or at least a related pattern that allows a simple and fast evaluation, be generalized to other songs containing only passages with nonsense syllables? The first example belongs to the so-called "Gstanzln" (Austrian-Bavarian mocking songs): "*Da ob'n auf'm Bergal… ho la di ri di jo, ho la di jo, ho la di ri di jo, ho la di jo…*". The second one is a very old German folksong: "*Ein Vogel wollte Hochzeit halten….fi di ra la la , fi di ra la la, fi di ra la la la la…*"

We assumed that in the nonsense passages of such songs the highest pitch would coincide with an [i]. In order to test this assumption we investigated all songs in Dawidowicz's (1980) collection that are assigned to the category Austrian folksong and include passages with nonsense syllables (Figure 2).

The results: Almost all songs (24 of 26) showed, in their passages containing exclusively nonsense syllable, without any exception the expected coincidence of the vowel [i] with the highest pitch in melody. Interestingly, most of the songs containing passages with nonsense syllables can be localized in the Alpine regions of Austria.

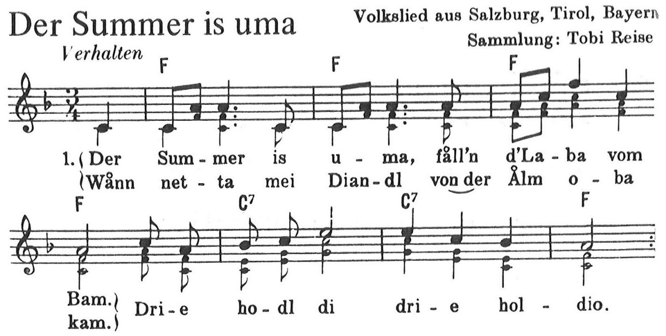

Figure 2. Two examples of folksongs (from Dawidowicz's collection, 1980, p. 94, 182) with passages of nonsense syllables: the vowel [i] in nonsense syllables coincides with the highest pitch in melody.

IV. CONCLUSION

If it is the aim to find a parsimonious explanation for the parallels reported between language and music, one may, first of all, address the mediating role of singing. If one assumes that vocal music shaped the most often only accompanying instrumental music, one may view a more or less archaic form of singing as a link between speech and *music* in a broader sense, i.e., in all its varieties. But the respective findings also seem to have some impact on the evolutionary models mentioned in our introduction.

Singing without words – as in the songs of animals, in "animal-like songs", in many utterances of our infants, in yodelling, in certain Arapaho songs (Nettl, 1956) – is cognitively less demanding than the use of a more or less arbitrary code. And it will well function without and before getting familiar with such a code. This reduces the plausibility of language as a precursor of music. For a closer inspection remain those two possibilities already favored in Jespersen (1895): Music either as "protolanguage" or both language and music as descendents of "half-musical" utterances.

But with respect to tonal modulation singing is more demanding than speech. For this and some other reasons it is more plausible that singing "prepared" the vocal tract for speech than the other way round.

A model accounting for these considerations, thus assuming a less specialized precursor with a more singsong- like tonality, not only allows for the correspondences between language and music in the repertoires, in phrase length and rhythmic structure of phrases. It also sheds some light on the key role of the vowel system and accentuates its role in language evolution. Vowels carry, as already mentioned, most of the syllables' sound or sonority in both singing and speech. During the evolution of speech that became the specialist for referential meaning (Brown, 2000), the sound of the less specialized tone-dominated precursor diminished. This, and the fact that the formant structure of the vowels remains the same whether the respective syllables are spoken or sung gives rise to some rather speculative assumptions:

(a) Maybe it is more than a mere coincidence that those high front vowels [i] or [e] that are easier to be sung in high tones (cf. Nettl, 1954), play a special role in phenomena of sound symbolism or phonosemantics, for instance, in the more or less universal tendency to be used in names for the sound or noise produced by the smaller exemplars of a category, and thus(?) also in the names for smaller exemplars of objects and in diminutive expressions. But a well-founded theory also requires further studies on commonalities regarding the sound structure, e.g. by integrating phenomena of sound symbolism and utterances of whistled languages which are intelligible for the reason (Meyer 2007) that they simulate acoustic cues carrying the melody of speech.

(b) The second speculation: Our considerations amount to the idea that the evolution of language was characterized by a gradual reduction of the tonal character of the common precursor of language and music. This is in line with and supports the conclusions of e.g. Morley (2003) and Masataka (2009, p.11). According to Masataka the prelinguistic system was neither identical with modern language nor with modern music, but "resembled music more closely than language."

To end up with a prospect of future research:

Again speculative, but in principle accessible for direct empirical testing is the question indicated in Section II whether cultures with a higher number of vowels also tend to use a higher number of pitches in musical melody.

REFERENCES

Brown, S. (2000). The "musilanguage" model of human evolution. In N.L. Wallin, B. Merker & S. Brown (eds), *The origins of music* (pp.271-300). Cambridge, MA: The MIT Press.

Burns E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (ed), *The psychology of music* (pp. 215-264). 2nd ed. San Diego: Academic Press.

Butcher, A., & Anderson, V. (2008). The vowels of Australian Aboriginal English. *Proceedings of Interspeech,* Bristol, Australia.

Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg (Ed.), *Universals of human language (*pp. 93-152). Stanford: Stanford University Press.

Darwin, Ch. (1871). *The descent of man and selection in relation to sex*. Retrieved August 24, 2004, from
http://www.darwin-literature.com/ The_Descent_Of_Man/21.html.

Dawidowicz, A. (1980) *Österreichisches Liederbuch. Komm, sing mit!* Innsbruck/Neu Rum: Musikverlag Helbling.

Deutsch, D. (2008). The ICMPC Series, and some current research on music perception and cognition. In K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, & M. Tsuzaki (Eds.), *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC10)*(pp.6-9).

Encyclopedia – Arapaho music (2009). Retrieved March 3, 2009, from
http:/www.experiencefestival.com/a/Arapaho_music/id/1916800

Fenk-Oczlon, G., & Fenk, A. (1999). Cognition, quantitative linguistics, and systemic typology. *Linguistic Typology*, 3-2, 151-177.

Fenk-Oczlon, G., & Fenk, A. (2002). The clausal structure of linguistic and pre-linguistic behavior. In T. Givón and B.F. Malle (Eds.), *The evolution of language out of pre-language* (pp.215-229). Amsterdam: John Benjamins.

Fenk-Oczlon, G., & Fenk, A. (2005). Cognitive constraints on the organization of language and music. In B.G. Bara, L. Barsalou & M. Bucciarelli (eds), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (p. 2476). Mahwah, NJ: Erlbaum.

Fenk-Oczlon, G., & Fenk, A. (forthcoming). Some parallels between language and music from a cognitive and evolutionary perspective. *Musicae Scientias, special issue "Music and Evolution*, 201-226.

Fitch, W. Tecumseh (2009). Musical protolanguage: Darwin's theory of language evolution revisited. Retrieved April 23, 2009, from
http://www.st-andrews.ac.uk/~wtsf/DarwinLanguageEvolution.html

Hughes D.W. (2000). No nonsense: the logic and power of acoustic-iconic mnemonic systems. *British Journal of Ethnomusicology, 9/ii,* 95-122.

Huron, D. (1996). The melodic arch in Western folksongs, *Computing in Musicology, 10,* 3-23.

Jespersen, O. (1983 [1895]). Progress in language. Amsterdam Classics in Linguistics 17. Amsterdam: John Benjamins.

Jespersen, O. (1922). *Language*. London: Allen.

Lauridsen, J. (1983). *Musical scales in Australian aboriginal songs: Structure and social implications.* Doctoral dissertation, University of Maryland.

Masataka, N. (2009). The origins of language and the evolution of music: a comparative perspective. *Physics of Life Reviews, 6,*11-22.

Meyer, E. A. (1896). Zur Tonbewegung des Vokals im gesprochenen und gesungenen Einzelwort. *Phonetische Studien*, *10*, 1-21.

Meyer, J. (2007). What does the typology of whistled forms of language teach us about prosody? In Book of Abstracts (p. 173), *7th International Conference of the Association for Linguistic Typology*, Paris, September 25-28.

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information: *Psychological Review, 63,* 81-97.

Mithen, S. (2005). *The singing Neanderthals.The Origins of Music, language, Mind, and Body*. London:Weidenfeld & Nicolson.

Morley, I. (2003). *The evolutionary origins and archaeology of music.* Doctoral Dissertation, University of Cambridge,

Nettl, B. (1954). Text-music relationships in Arapaho songs. *Southwestern Journal of Anthropology, 10*, 192-199.

Nettl, B. (1956). *Music in Primitive Culture*. Cambridge: Harvard University Press.

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. In N.L. Wallin, B. Merker and S. Brown (Eds.), *The origins of music* (pp.463 - 72). Cambridge, MA: The MIT Press.

Pommer, J. (1893). *252 Jodler und Juchezer*. Wien: Rebay & Robitschek.

Sachs, C. (1965). *The wellsprings of music*. New:York; MCGraw-Hill.

Sugiura, H., & Masataka, N. (1995). Temporal and acoustic flexibility in vocal exchanges of coo calls in Japanese macaques (Macaca fuscata). In E. Zimmermann, J. D.Newman & U. Jurgen (pp. 121-140). New York: Plenum.

The Arapaho Project (2009). Retrieved March 3, 2009, from
www.colorado.edu/csilw/arapahoproject/language/pronunciation.htm

Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge/Mass, London: MIT Press.