

STEPS TO A COMMON DESCRIPTION OF MELODY IN MUSIC AND SPEECH

Ernst Dombrowski

Department of Psychology, University of Kiel, Germany

ABSTRACT

Background and aims. Two modes of melody can be distinguished, the “plain” in speech and the “stylized” in music. Although these modes are separate they may be related to a unified melodic competence applying for both. Assuming such a competence it seems reasonable to try a common description of melody which is valid for music and speech.

Main contribution. A description of melodic structure will be represented applicable equally to music and speech materials. This description is mainly based on the Kiel Intonation Model (KIM). Essential traits of an intonation-based approach to melody will be worked out: (1) Melody arises from accent based patterns. (2) There are two forms of melody processing, local processing, directed to features of contour, and holistic processing, directed to the tonal reference space. (3) Intonational universals must be considered. A model of melodic syntax is set up that has a linguistic and a musical manifestation allowing a formalized mutual assignment of intonational and musical patterns.

Implications. The representation of melodic structure developed here can be implemented in a computer assisted analysis of musical signal data analogous to prosodic labelling. The intonation-based approach to music may be useful for performance analyses, for research on the declamation of texts in music, and for didactic purposes.

1. INITIAL CONDITIONS

Unified melodic competence for speech and music. There are two modes of melody, the one in speech and the one in music. These modes seem separate being connected with systems which differ in form and function. Nonetheless a unified competence can be assumed providing the basic conditions for melody production, melody perception, and the formation of melodic structure both in speech and music. The assumption of a competence of that kind requires a common description of melody despite its different modes - which will be suggested here.

Melody and vocal communication. The origins of melody can be seen in the pre-linguistic and pre-musical vocal behaviour of humans which is in turn related to the vocal behaviour of other species. Already 18th century aestheticians traced music and language back to sources in archaic vocalizations viewing these vocalizations as a kind of proto-melody (Rousseau 1984; Herder 1769). In this context, notions such as accent or accent of sentiment were used to designate the germ elements of melody - treating music and the musical properties of language in a very similar way. In some recent publications in phonetics and psychology the conception of a link between speech melody and

vocal communication has been reanimated (cf. Scherer 1991; Ohala 1984; Gussenhoven 2002) - and this may be extended to melody in music, too.

Five main aspects of speech melody can be derived from vocal communication and may also apply to music: (1) the use of high and low pitch following a universal *frequency code* (as proposed by Ohala 1984), conveying the communicative status of the speaker or addressee in a given situation, hence indicating speech act categories (e.g. statements and types of questions), and - more abstractly - signalling beginning, continuancy and closure, (2) the means employed for the highlighting and backgrounding of information, showing another use of pitch and effort which can be distinguished from the frequency code, (3) the vocal expression of emotive evaluations, (4) the nature of the basic melodic units - these units being communicative gestures of a characteristic tonal, dynamic, and temporal shape located in time, (5) the distinction of a preparation stage and an performance stage both required for a complete gestural unit (which can be simple or complex) - as is typical for voluntary motion.

Melodic gestures like this may have the character of a temporal gestalt - but they are clearly directional and therefore lacking the gestalt feature of reversability: E.g., an inversion of a pitch pattern would change a rise to a fall and thus alter melodic meaning.

Manfred Clynes stated comparable gestural units (*essentic forms*) in his theory of emotion (see Clynes and Nettheim 1982).

Melody: local gestures within a holistic frame. The relation between melody and vocal communication will be utilized here as a key to the understanding of melodic structure - both in speech and music. If melody is conceived as an arrangement of gestural acts, conveying linguistic or musical meaning, but derived from vocal communication, then two main characteristics of form in melody can be assumed, the first concerning the basic units of melody, the second with respect to the inclusion of basic units in large scale units.

(1) The basic elements of melody are *gestural patterns* of an (at first) *local scope* comprising a rising, falling or complex pitch contour, and a specific dynamic and temporal shape. These patterns occur at critical points in time, giving an impression of prominence or weight. They will here be called melodic accents and can be considered as the smallest melodic units of independent meaning (see Riemann 1903, p. 14).

(2) The formation or, conversely, identification of local melodic gestures is fundamental to melody processing. However, the description of melodic accents by their local features only will not be sufficient - because it does not show how a melodic accent is related to its surroundings: E.g., there may be strong or weak

melodic cohesion between the accent and its neighbouring accents, there may also be explicit breaks, preceding or following, and there may be an indication of beginning, continuation or closure expressed by the accentual pattern itself. To capture these aspects the local features of accents have to be completed by non-local ones. Combining local and non-local features will make clear in which way a single local pattern contributes to the formation of complete melodic units (being at least phrase-units) and to melodic structure as a whole. As a result, each melodic accent would be viewed from two perspectives: the *local* accent-level perspective and the integrating, i.e. *holistic* phrase-level perspective. Thus melodic accents can at the same time be individual and be parts or dominant elements of phrase-level units.

By parallel evaluation of local and non-local information some important properties of melodic form can be analyzed, above all, constructions with a preparing and a performing or completing stage. Constructions like this may be derived from formal principles which can already be found in single-accent patterns and which may be analogous to the organization of motion (or action). This dualistic principle has a counterpart in conceptual pairings as figure and ground, theme and rheme, beginning and cadence (Lerdahl and Jackendoff 1983), *Aufstellung* and *Beantwortung* (exposition and response, Riemann 1903) - which are of perceptual, linguistic, and musical importance (for these concepts see below).

Tonal sequence and tonal motion. A further indication of the nature of melodic structure and of melody processing is given by two terms often used as a definition of melody: *melody as tonal movement* and *melody as tonal sequence*.

The notion of *tonal motion* aims at a process, i.e., a continuous course forming an integral unit. This kind of unit may be conceived as gestural and therefore be related to gestural patterns (as rise, fall etc.). Above all, the local accentual melody can be conceived as movement, but motion can also be more extensive, i.e. holistic. The perception of movement is based on the process itself, not on pitch targets for example.

According to the notion of melody as a *tonal sequence*, melody can be subdivided into singular events which are characterized by one fixed pitch each. Without further consideration, the notion of tonal sequence does not make sense for the conceptualization of the process character of melody. Pitch events in the sense of "tones" can only be assigned an interpretation if there is a tonal space of reference in which tones are targets; and a space of reference will only be a space of reference by its tonal functions. Within speech melody, functional meaning is linked to the contrast of high and low pitch due to the *frequency code* mentioned above. I.e., by means of the link between rising and high vs. falling and low pitch to communicational function a global frame is established in which melodic structure is fitted. Within music, such a frame is formed by tonal functions which can be assigned to the steps of musical scales or cadential patterns etc. Thus the conception of melody as *tonal sequence* implies a look on melodic structure as a whole and cannot adequately be based on local processes.

By the distinction of terms like tonal motion or tonal sequence, two basic aspects of melody processing can be made clear which can also be confirmed by evidence from experimental research

in music psychology, cf. Dowling's (1978) concept of *scale and contour* and related studies by Edworthy (1985): According to Edworthy's studies, a manipulation of melodic contour can already be detected if it is inserted in the very beginning of a melodic stretch. For the detection of manipulations concerning musical intervals (the contour being preserved) more extended stretches of melody are required. As a reinterpretation of these results the following can be inferred: (1) Melody processing depends on a twofold competence. (2) There is a partial competence for the processing of melodic contour information, i.e. of pitch movements, already effectively operating on very short (i.e. local) stretches - and perhaps directed towards the detection of local accentual gestures. (3) There is another partial competence used for evaluating the contribution of local accentual patterns to large-scale units or for characterizing large-scale units for themselves. For this type of melody processing which can be called holistic, the scaling of pitches with respect to a stable space of reference becomes important - relating pitches to pitch functions.

2. RESULTING REQUIREMENTS FOR THE DESCRIPTION OF MELODIC STRUCTURE

As a conclusion from the considerations above, some basic principles of melody description can be proposed: (1) Melody should be described by an accent-based approach. Local melodic gestures (melodic accents) are the basic units of melody, being generative nuclei within the melodic phrase. (2) Melodic accents are characterized by local and holistic features, the local features marking properties of the local pattern, the non-local features marking concatenative and phrase-delimiting properties. Local features should be described in terms of contour (i.e. pitch movement), non-local features should include the relation of pitches or pitch movements to the tonal space established within a melody system, linguistic or musical. (3) Melodic accents should be systematically described by their form, showing a bipartite organization with a preparing and a performing stage, the latter starting at the critical point of time of the accent. (4) The bipartite organization should occur again at the level of large-scale units (e.g. phrases), these results also being derived from accent patterns.

A syntax of melody based on accentual gestures is outlined in [Figure 1](#). [Figure 2](#) shows the structuring of a single accentual pattern.

3. APPLICATION TO SPEECH MELODY

For an application to speech melody, the above principles of melodic structure must be adjusted to the conditions of the languages in question, which may, e.g., emphasize the accent level (as in German and English) or the phrase level (as in French). For languages as Japanese the moraic rhythm must be taken into account. Three current approaches to intonation, originally developed for Dutch, English, and German, will be compared to the requirements of melody description:

The *IPO approach* to intonation ('t Hart, Collier, and Cohen 1990) analyzes melody by induction from perception. Two basic

constituents of melody are posited as a result of the perceptual analysis: (1) a set of pitch movements, rising or falling, and (2) an inventory of declination lines, distinguishing a base line and one or more top lines. Pitch movements can be accent-related or not. If they are not accent-related they can be grouped in movements used to structure an utterance or to indicate non-declarativity, and finally in movements which just follow the declination lines. Thus there is an alternation of local and holistic stretches of contour, located in a reference space (cf. [Figure 3](#)).

The representation of melody by the *autosegmental-metrical (AM) approach* (Pierrehumbert 1980; for German see Grice and Baumann 2002; cf. Ladd 1996) starts with a phonological model. In this model, melody forms tone sequences consisting of categorical pitches only, to which downstep is added. Each pitch event is described in terms of local target tones, high (H) and low (L). But there are different kinds of tones, attributed to pitch accents or to phrase level phenomena (phrase accents and boundary tones). Thus the model allows for pitches which are local, and pitches which are holistic in the above sense - although both are described by pitch categories (H and L tones). In the AM representation of melody, pitch accents are nothing but local: There are no accents which are characterized by local and holistic features as well (cf. [Figure 4](#)).

The Kiel Intonation Model (KIM, Kohler 1991) describes intonation by integrating its phonetics, form, and function. According to the Kiel model, intonation phrases are built from accent contours (F0 peaks and valleys). Each accent contour receives a set of features some of which may be regarded as local and some as holistic in the above sense ([Figure 5](#)). Distinguishing local and global accentual features within KIM can be justified by the procedure of prosodic labelling proposed by the KIM-based PROLAB inventory (Peters and Kohler 2001): Accents receive labels for both their basic contour type and their concatenative or phrase-delimiting variation. Judgements on the (local) contour type are based on contour only; judgements on the concatenative and phrase-final (or phrase-initial) variations incorporate a pitch graduation referring to speaker's current range of pitch, mainly the base limit of voice being taken into account (Thus, in addition to the pitch contour, a reference space of pitch is used to evaluate the holistic aspects of contours). Cf. [Figure 6](#).

Because of its direction towards complete accent contours the Kiel model is also prepared to be supplemented by a *theory of accentual patterns*, dealing with the microstructure of accentual melody and its generative properties in macrostructure (this kind of conceptualizing accents is a prerequisite for the common description of melody intended in the present paper).

Thus, a review of the approaches to intonation considered here shows that they can be reread in the terms proposed for a universal description of melody. However, there are different degrees of correspondence. The universal conceptualization of melody outlined above is closest to the perspective of KIM; it was developed starting from this model.

Concerning intonational *large scale patterning*, the *bipartite organization* of gestural structure must be taken into account which was assumed above to be essential to the understanding of melody. Bipartite organization is a basic principle in speech melody contributing to the marking of theme and rheme,

background and focus, old and new information with the theme etc. often being associated with the preparing stage mentioned above and with the rheme etc. being linked to the performing stage. A basic case of large scale pattern formation in speech melody is the *hat pattern* - which was first described for Dutch by the IPO intonation research (Cohen and 't Hart 1967). The hat pattern is a symmetrical construction of at least two accents, combining a rising and (above all) high ending movement with a following falling one (see for an example the contour in [Figure 3](#); cf. Kohler 1991 for further formal description). For a functional interpretation of hat patterns see, e.g., Bolinger's thematic and rhematic accents (1986, p. 49) and again Kohler (1991).

4. APPLICATION TO MELODY IN MUSIC

The application of the above basic principles of melodic structure to music will be discussed here for western tonal music. The aspects and theoretical models to be mentioned will be related to those in the preceding section on speech melody.

A basic element of the IPO description of intonation is that the course of melody is held by declination lines. Trying a structural interpretation, these lines may represent centers of tonal gravity with a holistic domain these centers being characterized by a specific functional load (which is related to a high-low distinction). An interpretation of this kind would be paralleled by formal properties of *plainsong* as are characteristic for *Gregorian chant* and related musical idioms - which have a tenor and a final for each mode and sets of local melodic formula employed in initial, medial, and final position, and for the highlighting of words. Here, tenor and final form holistic centers of gravity in the pitch space (with the tenor as a point of reference for recitation and the final as point of reference for closure); melodic formula are inserted elements forming the contour. In plainchant, before the development of tonal harmonic music, the connection between functional tones (such as tenor and final) and the holistic high-low-distinction is evident from melodic contour.

The *generative approach in music theory* (as formulated by Lerdahl and Jackendoff 1983; *GTTM*) may be grouped together with the autosegmental-metrical school in intonational phonology both being based on the same conception of prosodic structure (founded on metrical hierarchies, mainly in terms of Liberman and Prince).

According to L&J (1983) the parallelism between music and language is one of time-span structure, i.e. of temporal organization: The temporal organization of music and language is characterized by hierarchically articulated relations of *head* and *elaboration of the head* (p. 330) the basic units being syllables in language and pitch events in music - which are both composed to units of progressively higher level of structure. L&J relate this kind of structure in time to structure in motor activities; they also refer to the notion of movements being preceded by preparatory movements. The principle of temporal organization stated by L&J is valid, e.g., for their description of a musical phrase: Phrases comprise a structural beginning (b) and a (nucleus-like) cadence (c) forming an "arc of tonal motion" (p. 31) ([Figure 7](#)).

But L&J (1983) do not relate linguistic and musical pitch and therefore exclude the most important issue of a possible kinship

of music and language. In another publication Lerdahl describes pitch categories, harmony etc. as “exclusively musical structures” (2001, p. 352) restricting the parallelism between melody and intonation to melodic contour - as expressed in units of high or low tones (in terms of the AM theory).

In *Riemann's approach* (e.g. 1884, 1903) the basic units of melody are motives, located at centres of metrical weight and covering one measure (*Taktmotive*). Motives are understood as gestural units comparable to vocal gestures, mimic gestures, gestures of the hand etc. Although consisting of level pitches they are perceived as continuous and are characterized by a defined dynamic shape. In accord with this, Riemann rejects the hierarchical decomposition of measures as applied by Hauptmann (Riemann 1884). Longer stretches of melody are formed by concatenation of local melodic gestures which are for themselves integral units. Cf. [Figure 8](#).

Riemann's approach is accent-directed in the above sense (rather than is true of GTTM). Moreover, R.'s view shows interesting parallels to the perspective which is now taken by KIM in intonational research: In particular, melody is set up by an arrangement of local contours which are integral. However, it is not clear whether musical motives can only consist of one or of more than one accent contours in the sense of KIM and whether they are accents or phrases - again according to KIM.

Riemann's view also gives ideas for intonational analogies in melodic large scale patterning: He posits a distinction of *Aufstellung* and *Beantwortung* which is already present within single motifs (forming an upbeat-downbeat relation) and which reoccurs at higher levels. This principle which can also be gathered from L&J's GTTM may be a musical counterpart for intonational theme-rheme organization. But again there is not yet a systematic account of the melodic patterns which convey cohesion within bipartite structure in music.

5. PLAIN AND STYLIZED MELODY

The above paragraphs were to show that melody in music and melody in speech are parallel in many respects. However, what makes them different is stylization being typical of music and exceptional in speech. The stylized mode of melody, as opposed to the plain mode, is marked by three features: by local level pitches, by fixed global pitch relations (i.e. defined intervals and thus a standardization of the tonal reference space), and by (usually) extended durations including fixed durational proportions. Stylization is linked to specific functional contexts which is already true of stylized intonation in speech (Ladd 1983); here level stylization is, e.g., used in call-contours. Stylized systems of melodic structure may develop in narrative situations, in religious contexts (liturgy), in infant directed communication (nursery songs), in dance and other group activities. Thus stylization brings out a focal situation of a situational background.

Level-stylization as a means of intonation research was already used by Isacenko and Schädlich (1970).

6 MUTUAL ASSIGNMENT OF MELODIC MODES

If there is a common description of melody applicable to both speech and music, there must be a set of rules or transformations enabling the mutual assignment of the plain and stylized modes of melody. In an accent-based approach these rules would start with accent units correspondently. As a precondition, a concept of accent microstructure is required giving a structural template which is independent of the accent's melodic contents. This template will provide the frame for the description of local contour and holistic tonal information within each accentual unit. Three parts of the accent-related pitch course are recorded: (1) the preparing movement within the preparing stage, (2) the onset movement and (3) the continuation movement, both within the performing stage (The latter movements can have a different semantic impact on the contour categories). This kind of description can be applied to both modes of melody, but, above all it can be used for an account of stylized accentual patterns - being a base for the rule-governed assignment of sets of stylized patterns to intonational accent categories (such as the peak and valley contours of KIM) (cf. [Figure 2](#)).

All contours consisting of level pitches (i.e. stepped contours) can be derived from the general expression in [Figure 9](#) - to which (holistic) information on tonal categories can be added (see the example in [Figure 10](#)). Relationships between plain linguistic and stylized musical contours can be systematically analysed by perception and production studies (Dombrowski, in preparation).

Three points must be taken into consideration in a parallel analysis of melody in music and speech: (1) In tonal music holistic aspects (in the sense of pitch functions) appear in two ways: as features of contour and as features of pitch category. (2) A perceptual equivalent transformation of plain to stylized contours (and vice versa) will often produce non-trivial results - which is interesting for the understanding of the nature of melody in music and speech - see for an example [Figure 11](#) where a plain rise-fall is changed into a stylized rising step. (3) A common representation of melody in music and speech will only be valid if musical tonal functions (including harmony) can be traced back to linguistic pitch functions and accent configurations. A starting-point for this will be the *bipartite organization* of melody and the melodic accent configuration within *hat patterns*, the latter perhaps being related to the cadential TSDT schema in tonal music (Dombrowski, in prep.).

7. FINAL REMARKS

The parallel representation of the linguistic and the musical mode of melody which is outlined above results in a comprehensive conceptualization of melodic structure. This representation can also be implemented in a computer assisted analysis of musical signal data. A musical labelling system analogous to prosodic labelling has already been constructed and has been tested for read and sung speech data (Dombrowski, in prep.). The intonation-based approach to music may be useful for performance analyses, for research on the declamation of texts in music, and for didactic purposes.

8. REFERENCES

1. Adriaens, L. M. H. (1991). *Ein Modell deutscher Intonation*. PhD diss. Eindhoven: Techn. Univ. Eindhoven.
2. Clynes, M. and Nettheim, N. (1982). The living quality of music: Neurobiologic basis of communicating feeling. In M. Clynes (ed): *Music, mind, and brain: The neuropsychology of music* (pp. 47-82). New York: Plenum.
3. Cohen, R, and 'T Hart, J. (1967). On the anatomy of intonation. *Lingua*, 19, 177-192.
4. Dombrowski, E. (1995). Über strukturelle Gemeinsamkeiten zwischen sprachlichen und musikalischen Melodien. *Musikpsychologie*, 12, 110-133.
5. Dombrowski, E. (in preparation). *Melodie und Intonation*. PhD diss. Inst. of Psychology, Kiel University. Kiel.
6. Edworthy, J. (1985). Melodic contour and musical structure. In P. Howell, I. Cross and R. West (eds), *Musical Structure and Cognition* (pp. 169-188). London: Academic Press.
7. Grice, M., Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte*, 191, 267-298.
8. Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. *Proc. Speech Prosody*. Aix en Provence.
9. 'T Hart, J., Collier, R., and Cohen, A. (1990). *A perceptual study of intonation: an experimental phonetic approach*. Cambridge: Cambridge University Press.
10. Herder, J. G. (1990/1769). Viertes kritisches Wäldchen. In: R. Otto (ed), J. G. Herder, *Schriften zur Literatur*, vol. 2/1, (453-640). Berlin: Aufbau.
11. Kohler, K. J., Hrsg. (1991). *Studies in German intonation*. AIPUK, 25, Kiel.
12. Ladd, D. R. (1983). Phonological features of intonational peaks. *Language*, 59, 721-759.
13. Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
14. Lerdahl, F. and Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, Mass.: MIT.
15. Lerdahl, F. (2001). The sounds of poetry viewed as music. In Zatorre, R. Z., and Peretz, I., *The biological foundations of music* (pp. 337-354). New York: New York Academy of Science.
16. Ohala, J. (1984). An ethological perspective on common cross-language utilization of f0 of voice. *Phonetica*, 41, 1-16.
17. Peters, B. and Kohler, K. J. (2001). Trainingsmaterialien zur prosodischen Etikettierung mit dem Kieler Intonationsmodell KIM. Unpublished manuscript. Kiel University, IPDS. Kiel.
18. Pierrehumbert, J. B. (1980/1987). *The phonetics of English intonation*. Bloomington: IULC.
19. Riemann, H. (1884). *Musikalische Dynamik und Agogik*. Hamburg: Rahter.
20. Riemann, H. (1903). *System der musikalischen Rhythmik und Metrik*. Leipzig: Breitkopf und Härtel.
21. Rousseau, J.-J. (1984). *Musik und Sprache: Ausgewählte Schriften* Wilhelmshaven: Heinrichshofen.
22. Scherer, K. R. (1991). Emotion expression in speech and music. In J. Sundberg, L. Nord, and R. Carlson (eds), *Music, language, speech, and brain* (pp. 146-156). Basingstoke: Macmillan.