Critical comparison of acoustical and perceptual theories of the origin of musical scales

Richard Parncutt, Department of Musicology, University of Graz

Invited presentation at INTERNATIONAL SYMPOSIUM ON MUSICAL ACOUSTICS (ISMA2001) "Musical Sounds from Past Millennia: The Preservation and Promotion of our Musical Acoustic Heritage", 10 -14 September, 2001, Perugia, ITALY

How can different theories of the origins of musical scales be weighed up against each other? What is the relative importance of different performance conditions, contexts, and limitations – acoustical, physiological, perceptual, cognitive, social? How can we hope to systematically describe the development of unrecorded monophonic pitch-time patterns and associated culturally unique (arbitrary) meanings? Answers to these questions are necessarily speculative, but on the basis of recent research, they can increasingly be grounded in scientific evidence. In the following I offer a possible account that focuses on perceptual theory and data.

Pitch-time patterns in infant-adult interactions. Infant-directed speech incorporates a vocabulary of contours or inflexions with specific emotional and social meanings. These inflexions may be represented as curves in pitch-time space (Papousek, Papousek & Symmes, 1991) and are to some extent cross-cultural. The ability to recognize them may develop by exposure to the mother's voice before birth (DeCasper & Fifer, 1980). The ability to communicate on this level may ultimately explain why melodies have meaning, and why humans are motivated to perform and attend to melody (Trevarthen & Deliège, 1999-2000).

Categorical perception. Speech sounds are perceived categorically, and the categories may be regarded as the perceptual basis of linguistic meaning (Liberman, Harris, Hoffman, & Griffith, 1957). Similarly, timbre space is perceptually divided into regions corresponding to vowels and sound sources (Grey, 1977). Categorical perception also underlies the division of the musical pitch continuum into scale steps (Burns, 1999). In both speech and music, categorical perception facilities (or enables) transmission, processing and storage of information (Terhardt, 1991).

Octave, fifth and fourth intervals. These "perfect" intervals are important for both western and non-western musics (Burns, 1999; Sachs, 1943). In a perceptual approach to the origins of scales, it is necessary to distinguish melodic (successive) from harmonic (simultaneous) intervals. Melodic intervals may have emerged earlier than harmonic, since everyday (and infant-directed) speech may, when perceived categorically, suggest melodic intervals. In any case, melody is more prevalent than harmony in world musics. The prehistoric emergence of melodic 8ve/5th/4ths cannot be *directly* explained by Helmholtz's (1863) theory of frequencies in common, because those frequencies are not consciously perceived by the listener; normally, one perceives only one pitch in a tone, corresponding to the fundamental. This conceptual problem may be solved by converting the physical spectrum into a perceptual spectrum of *possible pitch percepts*, weighted by their saliences or probabilities of being perceived (Terhardt, Stoll & Seewann, 1982b; Parncutt, 1989). On this basis, one can predict the tonal affinity or pitch commonality of each interval and compare these with similarity ratings of successive tones, or with the prevalence of melodic intervals in musical scores. Parncutt (1989) demonstrated the perceptibility of pitch commonality (as opposed to effects of familiarity with harmonic intervals and with tonal music) by comparing data on the perceived similarity of successively presented pure and complex tones, heard by musicians and non-musicians.

Pitch commonality is a possible, but not necessary, basis for the prehistoric emergence of 8ve/5th/4ths in musical scales. Another possibility is that *harmonic* 8ve/5th/4ths chronologically preceded their *melodic* (arpeggiated) versions. These harmonic intervals may have been preferred for their low roughness (Helmholtz, 1863; Hutchinson & Knopoff, 1978) or high perceptual fusion (Stumpf, 1883, 1890; DeWitt & Crowder, 1987).

Memory duration. A scale can only be established if its pitches can be held in memory for at least the duration of, say, a phrase or strophe. Rakowski and Morawska-Büngeler (1987) reported that a musician without absolute pitch could retain a single pitch (± 30 cents s.d.) for 5 minutes (see also Rakowski, 1996).

What about frequency ratios? Octaves, 5th and 4ths correspond *approximately* to the frequency ratios 1:2, 2:3 and 3:4. But the ear is remarkably insensitive to frequency ratios between simultaneous and successive pure tones (Allen, 1967; Plomp, 1967; Plomp & Levelt, 1965). Western musical intervals are perceived *linearly* and *categorically*, and intervals are defined by the center and boundaries of the category – not by ratios such as 5:4 or 81:64. Intervals can vary in size by up to a semitone (e.g., a major 3rd ranges from 350 to 450 cents: Burns, 1999). Typical intonations deviate systematically from frequency ratios (major 3rds larger than 4:5, 8ves larger than 2:1) and the size of the deviation depends on register (Rosner, 1999). The exact size of a performed interval is the result of a compromise between partially conflicting constraints (Terhardt, 1974a) such as roughness, temporal context, musical style, emotion, and melodic emphasis: intonation thus depends not only on sensitivity to the musical surface but also on cultural knowledge (Burns, 1999). Frequency ratios do not *directly* affect or determine intonation; instead, pure intonation minimizes *roughness* between harmonic complex tones (Hagerman & Sundberg, 1980; cf. Mathews & Pierce, 1980) – but most intonation is closer to equal temperament or Pythagorean than pure (Burns, 1999). These empirically based arguments cast doubt on ratio-based, abstract-mathematical theories of the nature and origins of scales.

A speculative, perceptually oriented history of Western scales

The following sequence is necessary speculative, incomplete and ethnocentric. Each point in the list is assumed to overlap temporally with adjacent points, so that several different points are active at any given historic (or prehistoric) time.

1. Unaccompanied melody (prehistoric)

(a) *Categorical pitch perception.* Music separates from speech as the pitch continuum is divided, in production (singing) and perception, into musically meaningful categories or *scale steps*. The archeological record (bone pipes) suggests that this occurred more than 40,000 years ago (Lawson, Cross, Scarre, & Hills, 1998).

(b) *Optimal melodic step size.* The most common interval between adjacent scale steps in world musics is about two semitones (P. G. Vos & Troost, 1989), and the smallest about one semitone. This may be explained as a trade-off between: promotion of melodic fusion (Bregman, 1990; Noorden, 1975), favoring small intervals (steps); limited discriminability in perception and accuracy in production (singing), setting a lower limit for interval size; and cognitive limitations of memory, which may set an upper limit on the number of scale steps (Burns, 1999; Eberlein, 1994; Rakowski, 1997). The just-noticeable frequency difference for successive pure tones is only 0.1 to 0.2%, or 2 to 3 cents (Wier, Jesteadt, & Green, 1977); but in musical contexts, the JND is increased by stimulus uncertainty, division of attention, short duration, low pitch salience, and masking (Watson, Kelly, and Wroton, 1976).

(c) *Melodic 8ve/5th/4ths*. As sensitivity to the tonal affinity (Terhardt, 1974a) or pitch commonality (Parncutt, 1989) of melodic 8ve/5th/4ths gradually develops, these intervals begin to be used deliberately.

(d) *Quintic pentatonic and heptatonic (diatonic) scales*. As intervals between scale steps are gradually tuned to 8ve/5th/4ths, the quintic pentatonic scale, corresponding to the black keys of the modern piano (but not equally tempered), emerges (cf. Arom & Fürniss, 2000; Sachs, 1943); and then the quintic heptatonic or diatonic (corresponding to the white keys). These cyclic patterns optimize the availability of 8ve/5th/4ths (point 1c; Burns, 1999; Gauldin, 1983; Huron, 1994) while simultaneously satisfying interval size constraints (point 1b).

(e) *Pitch range*. Musical pitch is initially confined to the ambitus of the human voice. A preference for this range is maintained even when musical instruments start to play outside it, because harmonic complex tones have the clearest pitch (highest virtual pitch salience) in this range (Huron, 1993; Terhardt, Stoll & Seewann, 1982a).

(f) *Prevalence, stability, anchor tones*. Scale steps that are consistently used more often, or sustained for longer, tend to be perceived as stable reference points (Cuddy, 1997; Krumhansl, 1990), leading to the different church modes.

(g) *Secondary anchors*. The second-most-used and hence second-most-stable tone tends to lie an 8ve/5th/4th from the primary anchor (cf. point 1c; Burns, 1999; Gauldin, 1983). The resultant hierarchy of stability facilitates cognitive processes of memorization and recall (Deutsch, 1980).

2. Harmonic dyads (ancient civilizations)

(a) *Harmonic 8ve/5th/4ths*. Men, women and children, when singing with each other or with instrumental accompaniment, tend to produce parallel 8ve/5th/4ths, which optimize perceptual fusion and minimize roughness (see above).

(b) *Fine tuning*. Sensitivity to mistuning (J. Vos, 1982) is finer in harmonic than melodic intervals, due to the more salient cues of beats and roughness. This leads to more precise tuning of scale steps (cf. McClain, 1979).

(c) *Intervallic roots*. The difference in pitch salience between tones in *harmonic* 4th/5th relationships (Terhardt, 1974a) is transferred through musical practice and memory to *melodic* 4th/5th intervals, and hence to 4th/5th relationships between scale steps. Stability of scale steps is now determined not only by prevalence and duration but also by intervals: the upper tone of a 4th tends to be a tonal center (P. G. Vos, 1999). This eliminates the church mode on B.

3. Polyphony (Middle Ages)

(a) *Contrary motion*. Polyphony may be said to begin when voices move independently (Nettl, 1963), promoting perceptual segregation (Bregman, 1990). Fusion *within* melodic streams is promoted by stepwise motion.

(b) *Harmonic 3rd/6ths.* Harmonic 3rds and (major) 6ths result from stepwise contrary motion from one 8ve/5th/4th to another (Dahlhaus, 1967).

(c) *Harmonic 2nd/7ths*. These sonorities tend to be avoided due to their roughness and/or failure to fuse. But their perceived dissonance can be reduced by stepwise voice-leading (Botte, Drake, Brochard, & McAdams, 1997) and by sound-level differences between voices (Terhardt, 1974b).

(d) *Major and minor triads*. Of all 19 triad types in the chromatic scale (Rahn's 1980 T_n*-sets*), major and minor triads are favored because they simultaneously satisfy 2c (fusion and root implication of perfect 5th) and 3c (avoidance of 2nd/7ths). Major triads are more prevalent than minor (Eberlein, 1994), because their roots (and hence their tonal function) are clearer (Parncutt, 1988, 1993; Terhardt et al., 1982a), and they fuse or blend more easily.

(e) *Root position and inversions*. Root-position triads are favored (Eberlein, 1994), again due to their clearer roots and superior tonal fusion (Parncutt, 1988). The same principles explain the relative rarity of 2nd-inversion triads.

4. Harmonic tonality (Renaissance)

(a) *Chroma salience profiles*. Musicians become familiar with the prevalence (probability) distributions of diatonic and chromatic tones following triads in musical contexts. In this way, they indirectly internalize the chroma (or pitch-class) salience profiles of major and minor triads (Parncutt, 1988).

(b) *Major and minor scales*. The practice of *musica ficta* – whereby for example lower neighbor tones are sharpened to produce leading tones, and (melodic and harmonic) tritones are chromatically inflected to produce perfect 5th/4ths (Bent, 1984; Judd, 1998; Randel, 1986) – blurs distinctions between church modes and leads to the consolidation of the modern major and minor scales. For example, sharpening the leading tone transforms Dorian into rising melodic minor, and Mixolydian into major (Lester, 1978; Werckmeister, 1687).

(c) *Triads as final sonorities*. As major triads in root position become familiar, they are perceived as consonant, and replace the 5-8 sonority at phrase endings. Major final triads are preferred over minor (*musica ficta; tierce de Picardie*).

(d) *Key profiles*. Chroma prevalence distributions of tonal passages (Knopoff and Hutchinson, 1983; Krumhansl, 1990) increasingly resemble the chroma salience profiles of final triads, strengthening the implication-realization effect (Meyer, 1956; Narmour, 1990) at the final cadence (Parncutt, in preparation), and accounting for the key profiles of Krumhansl and Kessler (1982).

The preceding account highlights the ambiguity of the term *tonality* and the extended historical period of its evolution (Dahlhaus, 1967; Eberlein, 1984). Already at point 1a it is possible to speak of tonality in the broad sense of *scale*, just as the early-19th-Century concept of *tonalité* referred primarily to (undifferentiated) pitches and intervals (Simms, 1975). A more specific definition is implied by 1f, according to which tones are perceived relative to a tonal center. The dominant-tonic relationship (1g, 2c) and triads (3d) are characteristic of major-minor or harmonic tonality (cf. Dahlhaus, 1967; Riemann, 1877). The system comes of age when triads not only act as tonics (4c) but also determine chroma prevalence (4d).

References

Allen, D. (1967). Octave discriminability of musical and non-musical subjects. *Psychonomic Science, 7*, 421-422.

Arom, S., & Fürniss, S. (Eds.) (2000). L'Afrique et l'Europe médiévale : la théorie du pentatonisme revue à travers les systèmes africains de tradition orale (Discussion Forum 1). *Musicae Scientiae*.

Bent, M. (1984). Diatonic *ficta*. *Early Music History, 4*, 1-48.

Botte, M.-C., Drake, C., Brochard, R., & McAdams, S. (1997). Perceptual attenuation of non-focused auditory streams. *Perception and Psychophysics, 59,* 419-425.

Bregman, A.S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

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

Cuddy, L. L. (1997). Tonal relations. In I. Deliège & J. A. Sloboda (Eds.), *Perception and cognition of music* (pp. 329-352). Hove, East Sussex: Psychology Press.

Dahlhaus, C. (1967). *Untersuchungen über die Entstehung der harmonischen Tonalität.* Kassel: Bärenreiter.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mother's voice. *Science*, *208*, 1174–1176.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics, 28*, 381-389.

DeWitt, L. A. & Crowder, R. G. (1987). Tonal fusion of consonant musical intervals: The Oomph in Stumpf. *Perception & Psychophysics, 41*, 73-84.

Eberlein, R. (1994). *Die Entstehung der tonalen Klangsyntax*. Frankfurt: Lang.

Gauldin, R. (1983). The cycle-7 complex: Relations of diatonic set theory to the evolution of ancient tonal systems. *Music Theory Spectrum, 5*, 39-55.

Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America, 61*, 1270-1277.

Hagerman, B., & Sundberg, J. (1980). Fundamental frequency adjustment in barbershop singing. *Journal of Research in Singing, 4*, 3-17.

Helmholtz, H.L.F.von (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (1st. ed.). Vieweg, Braunschweig.

Huron, D. (1993). A derivation of the rules of voice-leading from perceptual principles. Abstract: *Journal of the Acoustical Society of America, 93*, S2362.

Huron, D. (1994). Interval-class content in equally-tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Perception, 11,* 289-305.

Hutchinson, W., & Knopoff, L. (1978). The acoustic component of Western consonance. *Interface, 7,* 1-29.

Judd, C. C. (Ed., 1998). *Tonal structures in early music*. New York: Garland.

Knopoff, L., & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. *Journal of Music Theory, 27*, 75-97.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*, 334-368.

Lawson, G., Cross, I., Scarre, C., & Hills, C. (1998). Mounds, megaliths, music and mind. *Archaeological Review from Cambridge, 15* (1), 111-134. Cited in Cross, I. (1999), Is music the most important thing we ever did? In Suk Won Yi (Ed.), *Music, mind, and science* (pp. 10-39). Seoul: Seoul National University Press.

Lester, J. (1978). The recognition of major and minor keys in German theory: 1680-1730. *Journal of Music Theory, 22,* 65-104.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. S. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54,* 358–368.

Meyer, L. B. (1956). *Emotion and meaning in music*. University of Chicago Press.

McClain, E. G. (1979). Chinese tunings in late antiquity. *Ethnomusicology, 23*, 205-224.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.

Nettl, B. (1963). Notes on the concept and classification of polyphony. In A. A. Abert & W. Pfannkuch (Eds.), *Festschrift Friedrich Blume* (pp. 247-251). Kassel: Bärenreiter.

Noorden, L. van (1975). *Temporal coherence in the perception of tone sequences*. Eindhoven, NL: Institute for Perception Research.

Papousek, M., Papousek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development, 14*, 415-440.

Parncutt, R. (1988). Revision of Terhardt's psychoacoustical model of the root(s) of a musical chord. *Music Perception, 6*, 65–94.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer.

Mathews, M. V. & Pierce, J. (1980). Harmony and nonharmonic partials. *Journal of the Acoustical Society of America, 68,* 1252-1257.

Plomp, R. (1967). Beats of mistuned consonances. *Journal of the Acoustical Society of America, 42*, 462-473.

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America, 38*, 548-560.

Rakowski, A., & Morawska-Büngeler, M. (1987). In search of the criteria for absolute pitch. *Archives of Acoustics, 12*, 75-87.

Rakowski, A. (1996). Short-term memory for pitch. In H. Fastl, S. Kuwano, & A. Schick (Eds.), *Recent trends in hearing research* (pp. 99-129). Oldenburg: Bibliotheks- und Informationssystem der Universität.

Rakowski, A. (1997). The phonological system of musical language in the domain of pitch. In J. Steszewski & M. Jablonski (Eds.), *Interdisciplinary Studies in Musiology* (Vol. 3, pp. 211-221). Poznan, PL: Poznan Society for Advancement of Arts and Sciences (PTPN).

Rahn, J. (1980). *Basic atonal theory.* Schirmer, New York.

Randel, D. M. (1986). *The new Harvard dictionary of music*. Cambridge. Mass.: Harvard University Press.

Riemann, H. (1877). *Musikalische Syntaxis.* Leipzig: Breitkopf & Härtel.

Rosner, B. S. (1999). Stretching and compression in the perception of musical intervals. *Music Perception, 17*, 101-114.

Sachs, C. (1943). *The rise of music in the ancient world*. New York: Norton.

Simms, B. (1975). Choron, Fétis, and the theory of tonality. *Journal of Music Theory, 19*, 112-139.

Stumpf, C. (1883, 1890). *Tonpsychologie* (vols. 1 & 2). Leipzig: Hirzel.

Terhardt, E. (1974a). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America, 55*, 1061-1069.

Terhardt, E. (1974b). On the perception of periodic sound fluctuations (roughness). *Acustica*, *30*, 201-213.

Terhardt, E. (1991). Music perception and sensory information acquisition: Relationships and low-level analogies. *Music Perception, 8*, 217-240.

Terhardt, E., Stoll, G., & Seewann, M. (1982a). Pitch of complex tonal signals according to virtual pitch theory: Tests, examples and predictions. *Journal of the Acoustical Society of America, 71*, 671-678.

Terhardt, E., Stoll, G., & Seewann, M. (1982b). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America, 71*, 679-688.

Trevarthen, C., & Deliège, I. (Eds.) (1999-2000). Rhythm, musical narrative, and origins of human communication (special issue). *Musicae Scientiae*.

Vos, J. (1982). The perception of pure and mistuned musical fifths and major thirds: Thresholds for discrimination, beats, and identification. *Perception & Psychophysics, 32*, 297-313.

Vos, P. G., & Troost, J. M. (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception, 6*, 383-396.

Vos, P. G. (1999). Key implications of ascending fourth and descending fifth openings. *Psychology of Music, 27*, 4-17.

Watson, C. S., Kelly, W. J, & Wroton, M. W. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. *Journal of Acoustical Society of America, 60*, 1176-1186.

Werckmeister, A. (1687). *Musicae mathematicae*. Frankfurt & Leipzig (Lester, 1978).

Wier, C. C., Jesteadt, W., & Green, D. M. (1977). Frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America, 61*, 178-184.