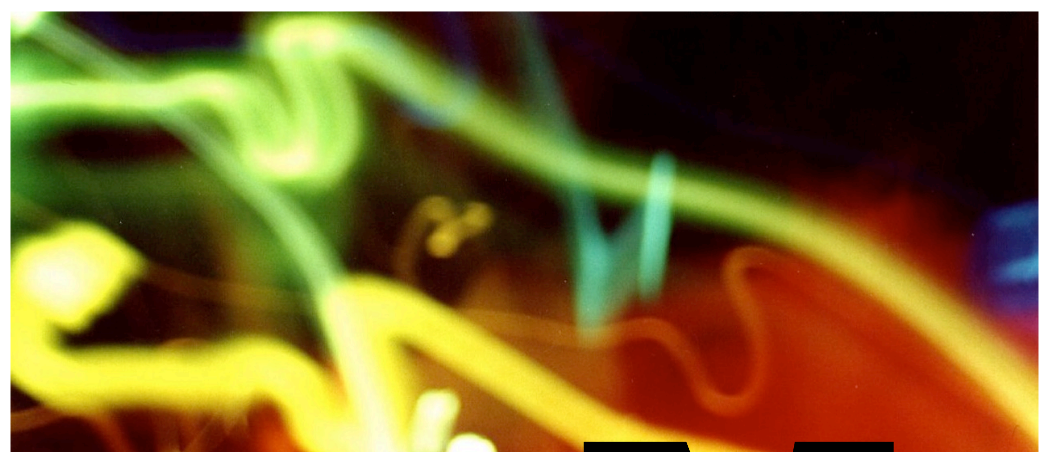
Sound Shapes

Rolf Inge Godøy
Department of Musicology, University of Oslo
e-mail: r.i.godoy@imv.uio.no
Sensing Music-Related Actions



fourMs

Music, Mind, Motion, Machines

www.fourms.uio.no

UNIVERSITY OF OSLO

Sound Shapes

- Hovedpoeng: Lære seg til å tenke selvstedig om lyd og lydattributter
- Ta utgangspunkt i subjektive inntrykk
- Danne begreper for lydattributter, gjerne finne opp nye dersom det ikke finnes fra før
- Overnfra-og-nedover ('top-down') og nedenfra-og-oppover ('bottom-up')
- Tiltagnede differensiering i attributtdimensjoner
- Tenke akser med verdier mellom min og max
- Knytte forbindelser til signalet
- Utfordringer i signalbehandling
- En mulig strategi for det hele: *Sound Shapes*

Sound Shapes

- Shape = what we experience as a multi-modal geometric unit
- Thinking shapes = general conceptual apparatus applicable to most music-related features, i.e. to timbre, rhythm, texture, expressivity, melody, modality, harmony, etc., as well as music-related movement, emotive and aesthetic features
- Let's have an example that demonstrates several instances of shapes:

Sound Shapes in beatboxing:

- Several sounds in succession, i.e. concatenations of ‘sonic objects’
- Each sonic object has a complex shape, i.e. several concurrent envelopes
- Vocal imitation of what are really not vocal sounds
- Attests to ‘motor theory’, i.e. motor simulation in auditory perception
- Attests to quite sophisticated motor control of vocal apparatus
- Suggests multimodal (sonic, motor, visual) shape-cognition, so let’s have a look at some elements:

1. Shapes: a unifying paradigm for musicology?

Notions of shape in recent musical thought

- Systematic musicology in the 21st century?
- Challenge of defining core issues of musicology in the face of increasing specialization
- Finding some conceptual scheme that applies across different approaches to musicology and for a renewal of musical aesthetics
- Shape is a candidate
- Shape well-known in various contexts, e.g. as melodic formulas in the renaissance and the baroque areas
- Shape one of the core issues of classic Gestalt theory

Shapes: a unifying paradigm for musicology

- One of the most extensive and well-founded shape-projects that of Schaeffer and co-workers (Schaeffer 1966)
- Schaeffer’s project still relevant and valid
- Schaeffer’s notion of the ‘sonic object’: basis for aesthetic considerations
- Accessible sources in English: the *Solfège* (with English text booklet), Chion/Dack, Godøy, etc., as well as indirectly in Smalley, Emerson, Landy, etc., and on the EARS web-site
- Developments of Schaeffer's ideas: the UST-project

UST

Les unités sémiotiques temporelles

- Chute (fall)
- Trajectoire inexorable (inexorable trajetory)
- Contracté-étendu (contracted-extended)
- Elan (dash)
- Etirement (stretching)
- En flottant (floating)
- Sans direction par divergence d'information (without direction by divergence of information
- Lourdeur (heaviness)
- Freinage (breaking)
- Obsessionnel (obsessed)
- Qui avance (which advances)
- Qui tourne (which turns)
- Qui veut démarrer (which will start)
- Sans direction par excès d'information (without direction by excess of information)
- Suspension – interrogation (suspension - interrogation)
- En suspension (suspended)
- Par vagues (by waves)
- Stationnaire (stationary)
- Sur l'erre (wandering)

Schaeffer’s and co-workers’ project:

- Focus on fragments of sound
- *Sillon fermé*: initially a practical necessity (but now living on in DJ scratching as ‘skip proof’)
- The typology: the overall shape of the sonic object
- The morphology: the internal features of the sonic object
- Typology and morphology based on listening, hence on subjective criteria
- Schaeffer claimed a non-linear relationship between subjectively experienced features and acoustics
- However, Schaeffer saw a long-term project of establishing correlations between percepts and signal (see e.g. Peeters and Deruty 2008 for a signal-based implementation)

Typological categories

- Sound categories suggested by Pierre Schaeffer (1966), and that correspond to biomechanically distinct action categories:
- Impulsive = discontinuous effort
- Sustained = continuous effort
- Iterative = rapid series of impulses, i.e. continuous effort but bouncing back and forth such as in a drum roll

Some morphological categories:

- Grain = Continuous movement across a rough surface
- Motion (“Allure”) = Slower fluctuations in harmonic content, in pitch, in loudness, etc.
- Schaeffer also suggested these categories may apply across different timbres and instruments

General point: Sound seems to be a good transducer of action information

Criteria for sonic objects

- Arbitrary cuttings
- Qualitative discontinuities
- Clearly determined by shapes e.g. as in the following sequence of quite different sounds:

Criteria of facture

Criteria of mass

Ordinary note	N	N'	N"
Complex note	X	X'	X"
Varied note	Y	Y'	Y"

- Extensive further qualification of criteria in the morphology, here summarized in the typo-morphological table:

		1	2	3	4	5	6	7	8	9
	Qualification (2-3) Evaluation (4-9) CRITERIA for musical perception	TYPES typo-morphological summary	CLASSES musical morphology	GENRES musical characterology	SPECIES (site and width of the dimensions in the musical field) PITCH SITE OF TESSITURA	WIDTH OF VARIANCE	INTENSITY SITE OF WEIGHT	WIDTH OF RELIEF	DURATION variations of emergence IMPACT	MODULE
1	MASS	TONIC type N COMPLEX X VARIABLE Y ANY W,K,T	1. Pure sound 2. Tonic 3. Tonic group 4. Stiped 5. Nodal group 6. Node 7. Fringe	Characteristic TEXTURES of mass	7 octaves x 12 = 84 steps HARMONIC COLOUR REGISTERS extreme deep very deep deep medium deep normal medium high high very high extreme high	HARMONIC INTERVAL COLOUR THICKNESS	WEIGHT OF HOMO-GENOUS MASS 1 ppp 2 pp 3 p 4 mf 5 f 6 ff 7 fff	PROFILE of the texture of a mass		(threshold of recognition for short sounds)
2	DYNAMICS	nil: homogenous H iterative Z weak weft N,X,T note shape N,X,N',X' impulsion N',X' cyclic Zk reiterated E accumulated A	Anamorphosis: SHOCKS RESONANCE Profiles: cresc. decresc. delta hollow biting Anamorphosis: flat	ATTACKS (dynamical timbre) 1. abrupt 2. steep 3. soft 4. flat pseudo attack biting 5. mild 6. pressed 7. nil			WEIGHT OF PRO-FILED MASS as function of its module 1 ppp 2 pp 3 p 4 mf 5 f 6 ff 7 fff	PROFILE MODULE weak medium strong	VARIATION OF PROFILE slow medium fast 1 2 3 4 5 6 7 8 9	SHORT SOUNDS MEASURED SOUNDS LONG SOUNDS
3	HAMONIC TIMBRE	either: GLOBAL TIMBRE or: secondary masses M1 M2 M3 ... timbre of the masses tm1 tm2 tm3 ...	(linked to the masses) NIL 1-7 TONIC 2 COMPLEX 6 CONTINUOUS 3-4 STRIPED 4-5	CHARACTER OF THE SOUNDING BODY hollow-filled round-sharp bright-dull etc.	COLOUR dark bright	WIDTH narrow wide 1 2 3 4	RICHNESS poor timbre rich timbre	density? volume? 1 2 3 4	variation: of width, of colour, of richness, from 1 to 9	(threshold of recognition for short sounds)
4	VARIATIONS MELODIC PROFILE	Unfolding / Profile / Anamorphosis Fluctation: N̄, X̄ / N, X / N', X' Evolution: Ȳ, T / Y, W / N', W' Modulation: Ḡ, P / G, M / G', K	(Only Y notes) podatus clivis toculus porrectus	character of profile: melodic pizz. trailing, etc.	or site of profile (see mass)	melodic variance weak medium strong	link of melodic profile with dynamic profile		slow medium fast 1 2 3 4 5 6 7 8 9	Partially or totally, se col. 3 start sustain decay
5	PROFILE OF MASS	Typological evolution Fluctation N /X or X/N Evolution Y/W or W/Y Modulation G/W or W/G	(Only thickness) dilatated delta slimming hollowing	characteristic evolution in mass in harmonic timbre	consequences for the tessitura or the colour (mass and harmonic timbre)	melodic variance weak medium strong	link of mass profile with dynamic profile		slow medium fast 1 2 3 4 5 6 7 8 9	Partially or totally, se col. 3 start sustain decay
6	SUSTAIN GRAIN	Pure or mixed by resonance scrubbing iteration	Trembling / Seething / Limpid rough / dull / smooth large / clear / fine	harmonic compact-harmonic compact compact-discontinuous discontinuous discontinuous-harmonic	GRAIN SEEN AS MASS OR TIMBRE colour of grain	thickness of grain	Relative weight GRAIN-MASS LINKED	Dynamic texture of grain weak medium strong	variation of grain width/speed from 1 to 9	tight adjusted loose 1 2 3 4 5 6 7 8 9
7	GAIT	Pure or mixed mechanical living natural	order / fluctuation / disorder 1 2 3 4 5 6 7 8 9	regular cyclic vibrato progressive irregular steep fall damped incident		variance of pitch of the gait weak medium strong	Relative weight allure/dynamic	Dynamic relief of gait weak medium strong	variation of gait width/speed from 1 to 9	tight adjusted loose 1 2 3 4 5 6 7 8 9

Summary of Schaeffer’s conceptual apparatus:

- Subjective notions of musical features the point of departure: name some feature and qualify it
- Increasing differentiation of features, sub-features, sub-sub-features, etc.
- Essential distinction between the *abstract* and the *concrete*:
- Abstract = symbolic values, e.g. pitches, scales
- Concrete = holistic emergent patterns, e.g. timbre, texture, various transients, nuances, etc.
- Shapes are multimodal, hence the fusion of Schaeffer's categories and action categories in our own work on *gestural-sonic objects*

And:

- Schaeffer's theory is about *sonic design*, hence also:
- An *analysis-by-synthesis* approach to aesthetics
- Schaeffer's theory had the ambition of being more universal than Western music theory
- Schaeffer saw non-western music as the greatest challenge for musical thought
- So, in our context: maybe inherited Western notions of musical aesthetics are too ethnocentric (similar to the music and emotion literature being dominated by western classical music categories)
- And: it seems we need concepts for musical features that are poorly, or not at all, thematized in Western thought

2. Embodied shapes: sound, action, vision, and amodal translations of shapes

Thinking Shapes

- Shapes everywhere, and needless to say, there are many thoughts on shape since antiquity
- In more recent times, Gestalt theory
- One of the most interesting projects, that of *morphodynmical* theory (e.g. Thom 1983, Petitot 1985, see Godøy 1993 for discussion)
- However, morphodynamical and related theories are mostly concerned with the emergent, self-organizing shapes; our main concern here is with shapes as a mental and experiential phenomenon, hence the embodied turn of shape cognition



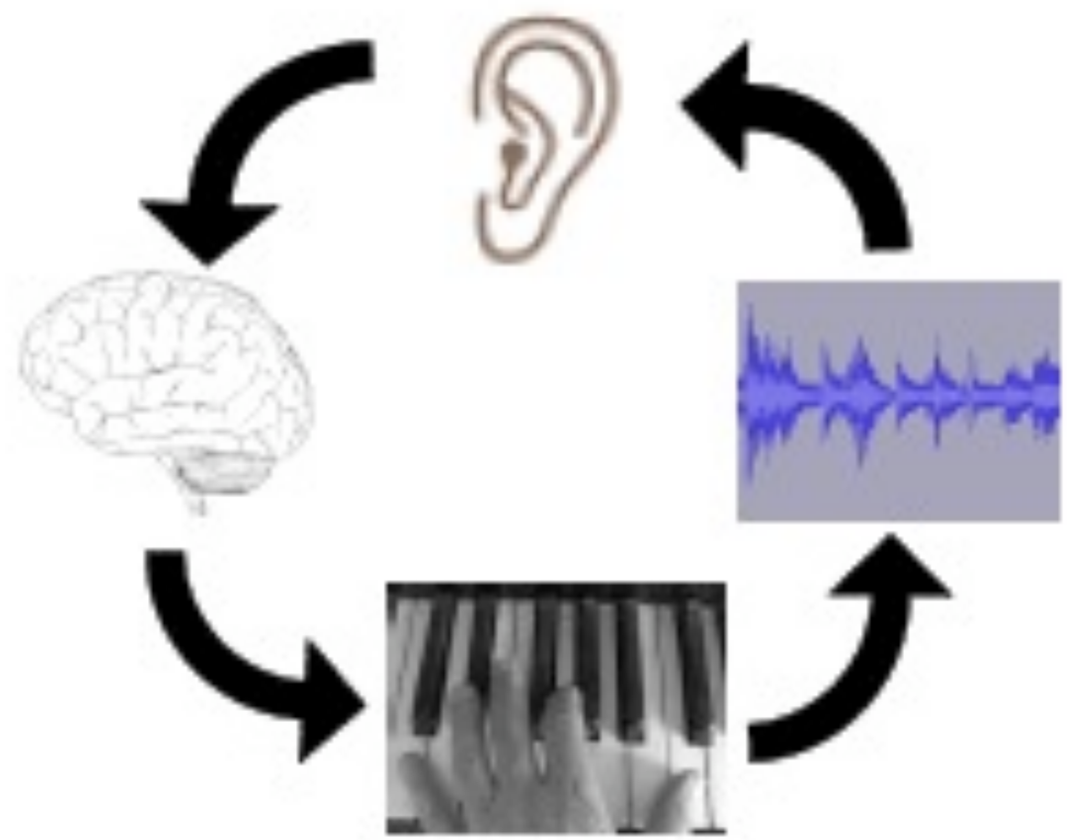

Embodied cognition:

- Large and fast growing literature on embodied cognition, and here only some of my favorites:
- Human movement involved in most perception and cognition (e.g. Berthoz 1997)
- Perception and cognition inseparable from bodily sensations (e.g. Gallese & Lakoff 2005)
- Interpreting phenomenology (Husserl, Merleau-Ponty) as embodied cognition (Gallese 2005)
- Suggestion: Movement and movement-related image schemata also basis for sound-cognition

Embodied shapes

- Converging research findings from various domains suggest ecological and multimodal elements at work in music perception and cognition, hence:
- An embodied turn (e.g. Leman 2008, Godøy & Leman 2010), and specifically, a motor theory perspective on music perception (Godøy 2003)

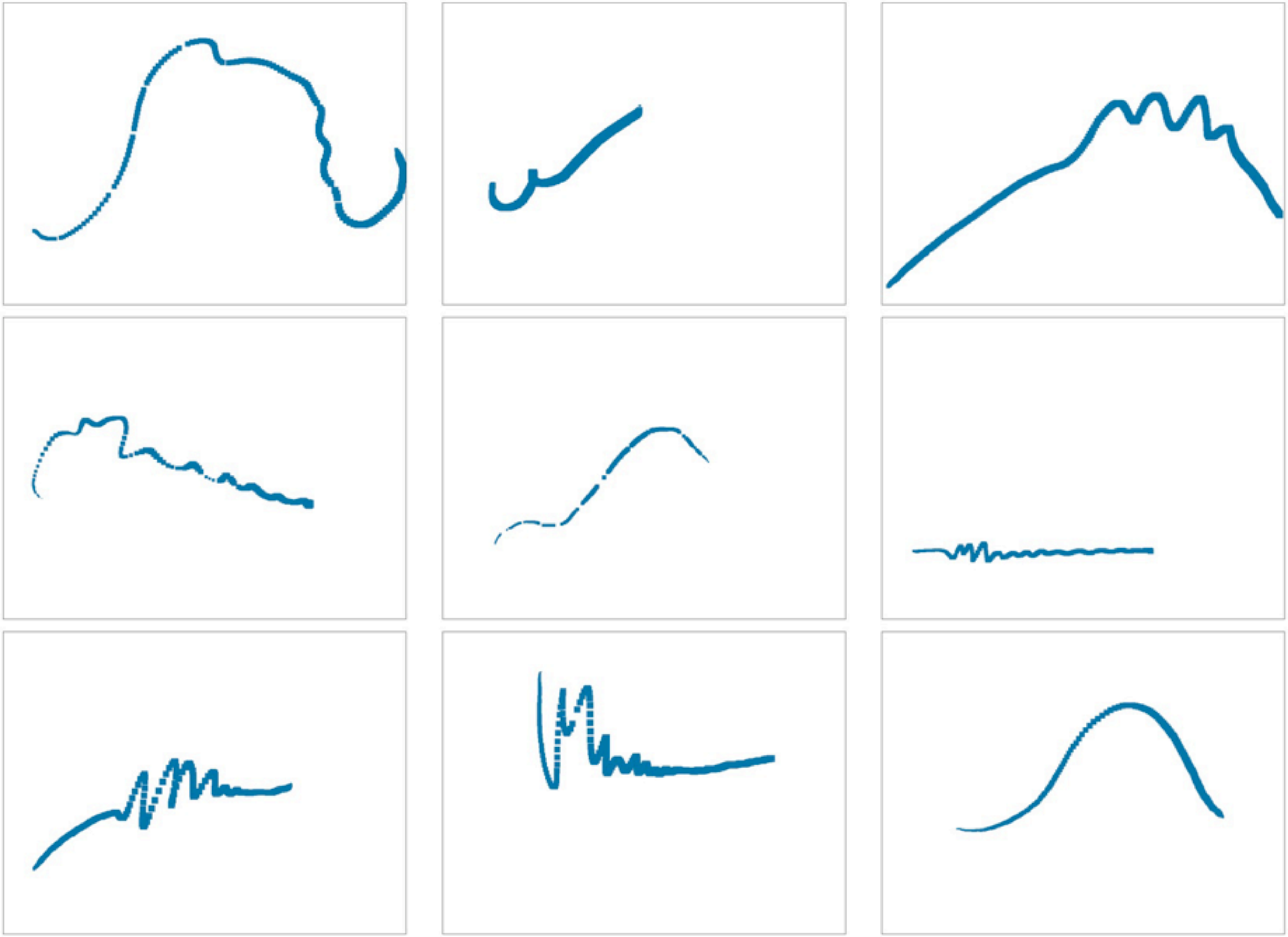
Motor theory as a perception-action cycle:



Crucial point of motor theories: variable acuity in motor images, i.e. not necessarily exact replication of sound-producing or sound-tracing actions; approximate, sketchy images are also valuable and interesting!

Motor theory of perception: Mental simulation and imitation of actions

- Motor theory initially in linguistics, but in recent years extended to several other domains
- Perception and cognition as a process of constant mental simulation
- Covert imitation of the actions of others and the assumed actions behind what we hear
- Neurocognitive research seems to confirm this incessant mental simulation/imitation going on
- Problem: Limited what we can see/observe
- Our various studies of music-related movement has tried to explore sound shapes related to musical performance:



3. Musical-epistemological issues: timescales and musical features

Epistemological issues

- Shapes everywhere, essential to clarify our focus:
- Timescales and features from micro to macro
- Schaeffer suggested that the sonic object timescale is most salient
- Good reason to be sceptical of inherited notions of large-scale forms, cf. Eitan & Granot 2008, Gerdingen & Perrott 2008, etc.
- But there are of course also essential features at more macro-levels, i.e. as dramaturgy/narrative and other effects
- Interesting to study macro-levels of music-related movement:

Timescales - a three-level model:

- *Sub-chunk level*: Continuous sound and movement below the chunk level of duration (i.e. below roughly 0.5 seconds)
- *Chunk level*: Holistically perceived fragments of sound and movement roughly in the 0.5 to 5 seconds range as with Pierre Schaeffer's *sonic objects*
- *Supra-chunk level*: Concatenations of chunks into larger scale units, i.e. into sections, movements, and whole works

Our present focus is on the chunk-level:

- Chunks are perceived and conceived as coherent units
- Chunks represent a significant time-scale in music, usually sufficient to perceive style, genre, mood, and most other features
- Regard individual tone-events/sound-events as included in chunks, hence as contextually smeared
- Regard chunks of musical sound as coinciding with action chunks

Phenomenological issues

- Husserl (1893): chunking by necessity: although sensations are sequentially unfolding (‘in time’), chunks also perceived and conceived ‘instantaneously’ in a ‘now-point’
- Thus: holistic perception of chunks, including the recent past (*retention*), the present, and the near future (*protention*) suggested by Husserl
- Incidentally: Husserl’s protention seems similar to present ideas on anticipatory cognition

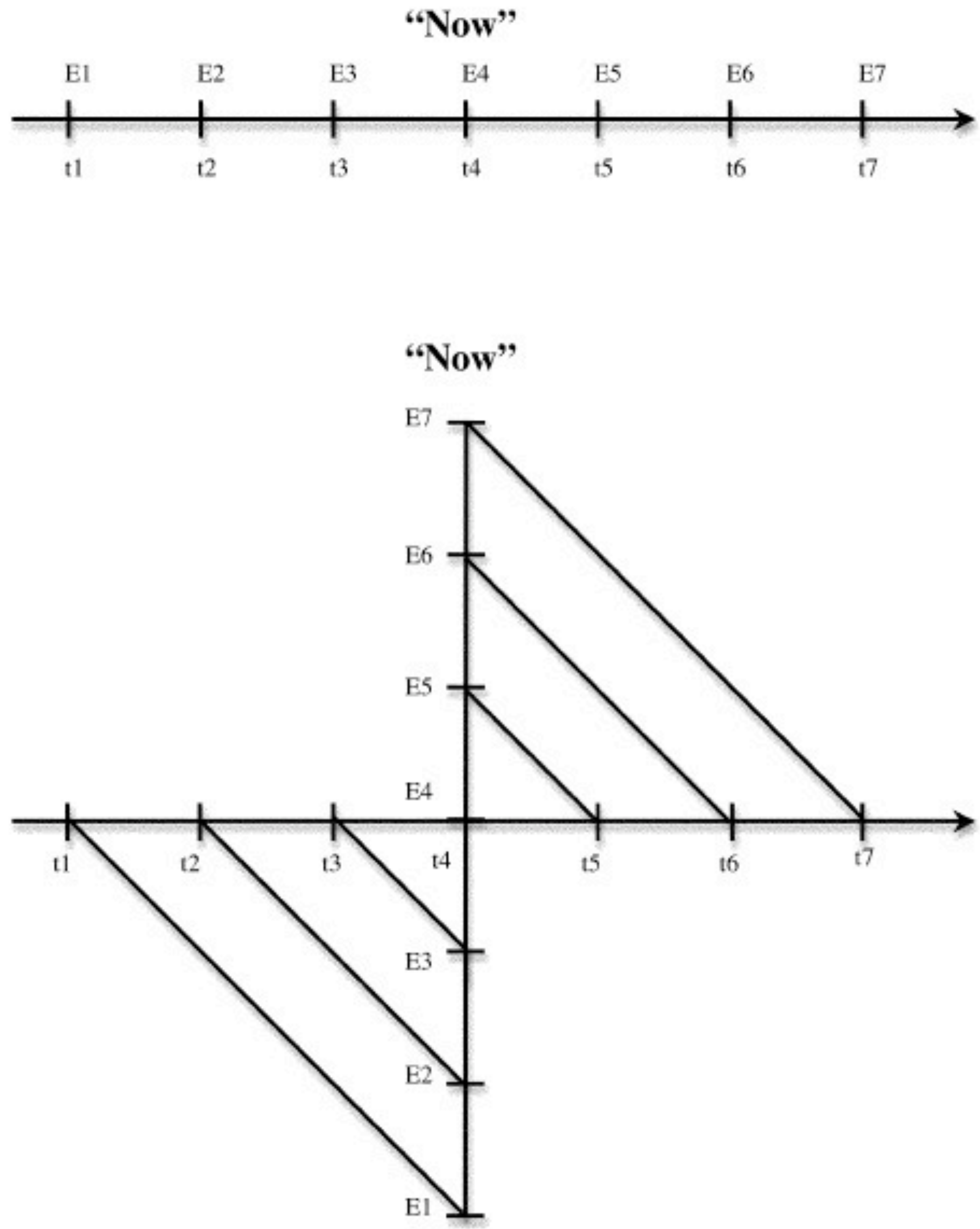


Figure 1: A depiction of the now-point based on Husserl's texts as well as on his various figures in (Husserl 1991). See main text for more information.

Holistic perception and conception of chunks:

- Enigma of how the sequential can be instantaneous
- Possible answer for sound: echoic memory
- Possible answer for movement: also a kind of short-term memory
- But also anticipation in motor cognition
- Various evidence for anticipation in motor control from Lashley to Rosenbaum (see Rosenbaum et al. 2007 for a lucid overview)
- When/how does an image of a chunk arise?
- How long does it last?

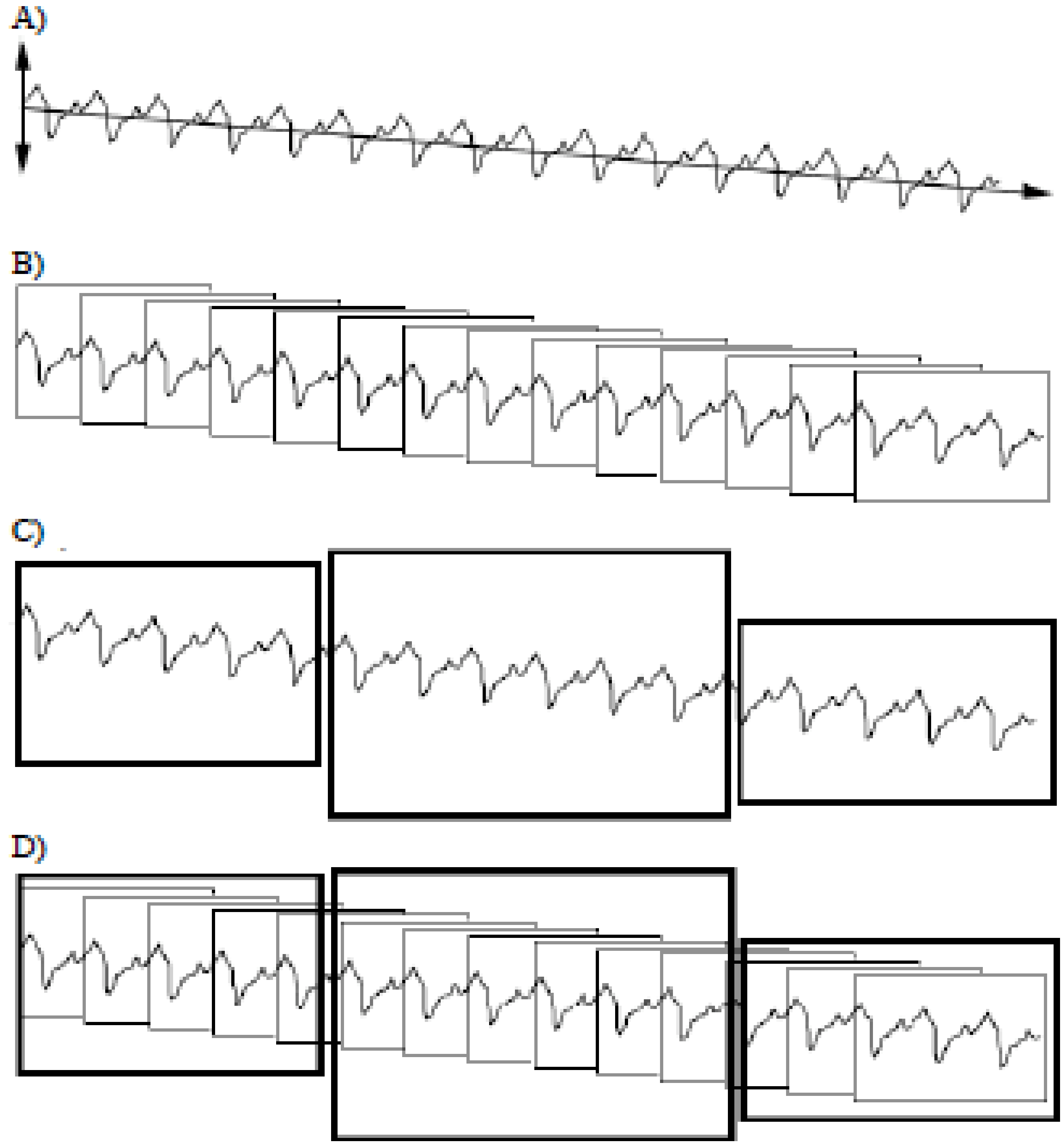


Figure 1. Given a continuous stream of sensory impressions, be that sound, vision, or motion, A), is there a more or less continuous updating of our awareness by a 'sliding window', B), or is there a more discontinuous updating of our awareness by disjunct chunks, C), or is there a combination of continuity and discontinuity in our awareness, D)?

Holistic perception and conception of chunks:

- Chunks work on a limited timescale, typically around 3 seconds (flips for longer durations, cf. Pöppel)
- Everyday actions demonstrate chunk size
- General features: movement is usually continuous, i.e. does not change abruptly, but is usually not very long, and is usually also goal-directed

Features of chunks in music:

- Rhythmic textural patterns: All kinds of rhythmical fragments, including cyclical patterns, also at work in entrainment
- Timbral contours: All kinds of changes over time, various transients, fluctuations, etc.
- Modal/tone-semantic and harmonic features
- Melodic contours: All kinds of melodic shapes
- Expressive features, i.e. timing, articulation, accents, phrasing, “feel”/”groove”, etc.
- Chunking both a perceptual phenomenon and an analytic strategy
- But what are the sources of chunking?

Sources of chunking:

- Exogenous sources of chunking: qualitative discontinuities and low-level gestalt features in the signal, as well as repeated patterns
- Endogenous sources of chunking: chunking imposed on the signal by our minds/bodies, c.f. cases of projection of metrical patterns onto pulses
- But also chunking by effort/biomechanical constraints (need for rests, posture changes, etc.) and by attention constraints
- Chunking due to anticipation, both biomechanical (preparatory movements) and motor control (goal-directed movement)
- Constraint based-chunking: *coarticulation*

Case study: coarticulation

- Coarticulation represents the convergence of shape-thinking with sound, action, vision and basic issues of time and continuity

Coarticulation

- Coarticulation = Fusion of micro-level actions and sounds into meso-level, holistically experienced chunks of actions and sounds, entailing a contextual smearing of the micro-level elements
- One advantage of coarticulation: Can account for the holistic perception, cognition, as well as motor control (anticipatory) of sound-action chunks
- The beauty of coarticulation: May work both forwards and backward in time, i.e. future events are colored by past events and past events are colored by future events

Coarticulation in various domains:

- Everyday tasks, e.g. reaching and lifting
- Animation
- Facial movements
- Fingerspelling
- Handwriting
- Music, but not well studied here
- Much studied in speech (see Hardcastle and Hewlett 1999 for an overview):

Coarticulation in speech

(from: http://person.sol.lu.se/SidneyWood/coart/coartint/coartina.html)

Coarticulating "happy"

For example, suppose you say the word happy:

Before you say anything, you will have moved your tongue into position for a

Then, while you are saying h, it will sound a bit like a

While you are saying a, you will also be closing your lips for pp

While your lips are together for pp, you will be moving your tongue to where you need it for y

Finally, while you are saying y, you will be opening your lips after pp

The whole word will usually be uttered in less than half a second

Principles of coarticulation:

- Otherwise singular events embedded in a context
- Past events influence present events, i.e. position and shape of effectors are determined by recent action
- Future events influence present events, i.e. position and shape of effectors are determined by preparation for future actions (anticipatory movements)
- Seems to be a biomechanical necessity
- Seems to be a motor control necessity, i.e. anticipation in motor control

Principles of coarticulation:

- coarticulation can be seen as an advantageous element: "…it is a blessing for us as behaving organisms. Think about a typist who could move only one finger at a time. Lacking the capacity for finger coarticulation, the person's typing speed would be very slow. Simultaneous movements of the fingers allow for rapid responding, just as concurrent movements of the tongue, lips and velum allow for rapid speech. Coarticulation is an effective method for increasing response speed given that individual effectors (body parts used for movement) may move relatively slowly." (Rosenbaum 1991, 15)

Principles of coarticulation:

- Basically: Body movement tends to be continuous, and also results of actions tend to be continuous (however sometimes very briefly)
- Can in some cases also be understood as a mass-spring phenomenon, i.e. as overlapping resonating events
- Has consequences for perception
- Contextual smearing in sound
- Contextual smearing in movement

Some studies of coarticulation in sound production:

- In piano playing: fingers move to optimal position before hitting key (Engel, Flanders, and Soechting 1997)
- In string playing: left hand fingers in place in position well before playing of tones (Wiesendanger, Baader and Kazennikov 2006) and contextual smearing of bowing movements (Rasamimanana and Bevilacqua 2008)
- In drumming: In some cases, a drummer may start to prepare an accented stoke several strokes in advance (Dahl 2004)

Metaphors

Metaphor theory:

- Great interest in metaphor theory since the 1980's in particular in so-called 'cognitive linguistics'
- Now commonly agreed hard to distinguish non-metaphorical and metaphorical language - try some 'metaphor avoidance' yourself!
- Metaphors are necessary and useful
- Metaphors are conceptual tools for structuring and communicating our sensations and thoughts to others
- Source domain and target domain
- Catachresis (Black 1962), e.g. 'orange' from fruit to color

Image Schemas:

- "Fundamental embodied cognitive structures generalized form recurring physical experiences, especially the experiences of our own bodies"
- Recommended reading: Johnson 1987, Lakoff 1987, etc.
- Convergence with recent neurophysiological research, highly recommended Gallese & Lakoff 2005
- Motor theory and image schemas: Our Musical Gestures Project
- Image Schemas have a certain degree of generality and even 'abstraction', applicable in many situations, e.g. our schemas for sound-production, effort and sound-shapes
- Often spatial relations in image schemas

Image Schemas and Music:

- Image schemas are conceptual phenomena, not necessarily unambiguously encoded in the sound, i.e. there is a cultural component here
- Think of various image schemas in different musical cultures, e.g. our 'high'-'low' schema for pitch is not universal

"Up" and "Down": Music and Gravity:

- Very prominent schema in western musical culture
- Applicable to many musical features - try to think of some examples!
- For a discussion of cultural vs. more universal elements in various music-related metaphors, see Eitan and Granot 2006

Various Spatial Metaphors in Music:

- 'Centrality in Music': Prominently spatial metaphor with several applications in music, and 'landmark' notion for pitch, harmony, motion, etc., and for particularly salient events, form, etc.
- 'Motion-Linkage-Causation': Goal-directed motion, as well as effects of events, cf. Sundberg & Friberg on runners retardation curves and ritardando in music
- 'Linearity: Paths and Goals': Leading towards events
- 'Containment: Inside and Outside': Several applications such as closure, but also style and all kinds of features

Metaphorical Extension:

- Various extensions and possibilities: Pretty much any feature in music that we may wish to focus on may be labeled by a metaphor and subsequently differentiated further (Godøy 1993)
- Assigning metaphorical labels is a way to focus our attention on some feature, hence, is actually also a research strategy
- In summary: Many metaphors in music-related discourse, visual, haptic, motor, etc., but most prominent is probably spatial
- Case study: Thomas Porcello's *Speaking of Sound*, where the student/ drum tuner (JM) asked the producer (DE) to describe 'the drum sound' DE wanted:

1. DE: But yeah it's like the main thing I look for is, is, you know just to make sure that they're gonna be singing out, that they have a lot of sustain to them. 'Cause what happens I find is um if they're, if you don't get, if you don't get a lot of sustain to them then what, they, the hit quickly comes and goes and doesn't really get heard in the track, so it's like that after-ring that'll make you sound powerful
2. JM: OK, yeah, yeah, alright
3. JM: So no pitch-bend?
4. DE: No
5. JM: Alright, cool, wide open sound
6. DE: Yeah
7. JM: No ... muffling
8. DE: Nah
9. JM: You want that?
10. DE: Nah
11. JM: No muffling, OK, good
12. DE: Lots of, lots of ring
13. JM: Lots of ring, no muffling, uh
14. DE: Yeah

Cross-modality

Our senses:

- Classic notion of 5 separate senses: sight, hearing, touch, smell, taste: a classification attributed to Aristotle
- Now more differentiated (listing more senses), but now also:
- Recognition of interdependence of senses, e.g. balance as composite of sight and vestibular sensations, and (in our case) hearing as influenced by sight and motion.
- Also *amodal* phenomena, i.e. mental schema that applies to several senses

Cross-modality in music:

- Music and action, cf. earlier lecture
- Consider other examples: Visual scenes, landscapes, shapes, etc. related to music
- Confluence of different modalities, e.g. in cinema
- On global, mood levels
- On more local, event synchrony levels, e.g. “Mickey mousing”
- Recommended reading: Michel Chion’s Audio-Vision

Derfor: Lyd og lydattributter som 'shapes':

- Lyd-relaterte handlinger in integrert del av lydbildet, dvs. lyd er ikke signal alene
- Lydproduserende gester del av lyden
- Lydakkompagnerende gester en del av lydperspesjonen
- Styringsgester en del av syntese/dsp
- Og (som vi skal se etterhvert), gester som en del av MIR

Lyd og lydattributter som 'shapes':

- Tidsdomene (bølgeform), og særlig:
- Fluktureringer på grain-nivå
- Allure-nivå
- Objekt-/chunk-nivå, typisk envelopper for enkeltlyder/-toner
- Og underdimensjoner av alle disse, dvs.
- Periodisitet
- Fluktueringer i periodisitet
- Onset-deteksjon (også i frekvensdomene)

Lyd og lydattributter som 'shapes':

- Frekvensdomene (spektrale attributter)
- Ulike FFT-bilder
- Andre representasjoner, f.eks. ulike typer waveletts
- Perseptuelle modeller (øremodeller)
- Spektral centroide
- osv., jfr. Peeters & Deruty 2008 (og Peeters 2004)
- osv.
- osv.
- Merk: Instrumentkarakteristika veldig viktige, men nå ha et helt universelt lydperspektiv

Lyd og lydattributter som 'shapes':

- Danne grunnlaget for Sonic Design vurderinger
- Danne grunnlaget for systematiske Analysis-by-synthesis utforskninger
- Danne grunnlaget for musikkanalyse generelt, og for sound-, instrumentasjons-/orkestrasjonsanalyse spesielt
- Danne grunnlaget for framførings- og ekspressivitetsanalyse
- Danne grunnlaget for MIR

Table 3.3 Examples of low-level features.

Feature	Description	Calculation
ZCR	Count (positive) zero crossings within N samples	$\sum_{k=0}^{N-2} x(k+1) \geq 0 \wedge x(k) < 0$
RMS	Root mean square amplitude calculated over N samples	$\sqrt{\frac{\sum_{k=0}^{N-1} x(k)^2}{N}}$
Max power	Maximum power in a block of N samples; often used in sample editor waveform displays when zoomed out	$\max_{k=0}^{N-1} x(k)^2$
Spectral centroid	Statistical measure over the spectrum	$\frac{\sum_{k=0}^{N/2-1} k\|X_m(k)\|^2}{\max(\sum_{k=0}^{N/2-1} \|X_m(k)\|^2, 1)}$
Spectral flux	Change of spectrum between frames	$\sum_{k=0}^{N/2-1} \|\|X_{m+1}(k)\|^2 - \|X_m(k)\|^2\|$
Spectral fall-off	The spectral envelope can be modeled by fitting a curve to the magnitude spectrum. Spectral fall-off fits a single line to model the typical drop in energy at higher frequencies in sound, as one helpful timbral indicator, but more complex models are available	Rodet and Schwarz [2007]
LPC coefficients	Linear predictive coding models the spectrum of the input with a source-filter model; it is a useful compression technique	Gold and Morgan [2000]; Rabiner and Juang [1993]; Makhoul [1975]
MFCCs	Mel-frequency cepstral coefficients; given a spectrum, the **cepstrum** approximates the principal components, and is a useful timbre descriptor; it also deconvolves (separates) an excitation and body response and gives some idea of pitch	Gold and Morgan [2000]; Logan [2000]; Roads [1996, pp. 514–8]

Table 3.4 Examples of higher-level features.

Feature	Description	Review references
Onset detection	Identifying the physical beginning of sound events	Bello *et al.* [2004]; Collins [2005]; Dixon [2006]
Pitch detection (monophonic)	Finding the fundamental frequency that would be selected by the human auditory system	de Cheveigné [2006]; Gómez *et al.* [2003]
Melody extraction	Transcription of a lead melody line, for example, as a sequence of discrete notes	Gómez *et al.* [2003]
Pitch detection (polyphonic)	The more general case of multiple simultaneous voices	Klapuri [2004]; de Cheveigné [2006]; Klapuri and Davy [2006]
Key and chord recognition	Detection of harmony	Gómez [2006]
Beat tracking and rhythm extraction	Determination of tempo, of beat locations and other metrical structure, and of rhythmic patterns	Gouyon and Meudic [2003]; Gouyon and Dixon [2005]
Instrument recognition	Timbral categorization	Herrera-Boyer *et al.* [2003]; Klapuri and Davy [2006]