## M03 • Pitch, harmony and counterpoint

ν(d) = ν_ref 2^((d−d_ref)/D); ν_i/ν_j = p/q; cents(i,j)=1200 $\log_2$(ν_i/ν_j).

Tune = (reference,interval system,continuous inflections,context); Mode = (intervals,gravity,phrase grammar,ornament,intonation).

h_k = (root,quality,extension,inversion,function,tension,duration,voicing,bass).

E_VL = Σ_i w_i|p_i,k+1−p_i,k| + λ_r E_range + λ_x E_cross + λ_p E_parallel.

m = (interval contour,rhythm,accent,register,gesture,semantic identity); $\mathscr{T}$_m = {transpose,invert,retrograde,augment,diminish,sequence,fragment,reharmonize,reorchestrate}.

E: frequency and cents relations. M: voice-leading penalties are tradition-dependent. Twelve-tone equal temperament is optional; raga, maqam and patet require their own phrase grammars. Independent lines, imitation, canon, fugue, suspensions and contrary motion are available operations, not guarantees of quality. Dissonance may resolve, prolong or remain.

## M04 • Time, groove and form

B(s)>0; b(s)=$\int_0^s$ B(u)/60 du; s_j=$b^{-1}$(b_j+δ_j); n_j=round(F_s s_j); |n_j/F_s−s_j|≤1/(2F_s).

Meter = (grouping,accent,cycle); Groove = (Cov(onset,velocity),swing,anticipation,drag,syncopation).

F = (motif→gesture→phrase→section→movement/album); D(s)=(level,density,register,brightness,roughness,space).

Rests ⊂ Events; Breath ⊂ Structure; p:q cycles require a declared common temporal reference.

The inverse beat map needs positive tempo and sufficient regularity. Additive meter, polymeter, tala, theka, sam, khali, clave, cascara and gong cycles keep their cultural context. Strophic, AABA, verse/chorus, blues, rondo, sonata, fugue, variations, through-composed, alap/jor/jhala plus gat, call-response and process forms remain distinct templates.

## M05 • Affect, meaning and asymmetry

l_j=(text,IPA,stress,syllables,breath,metaphor,subtext,persona,listener,story time,valence,arousal,agency).

a(s)∈$[0,1]^{12}$ = (grief,longing,awe,tenderness,anger,fear,serenity,hope,desire,nostalgia,ambiguity,resolve).

G_sem=(images,meanings,stakes,repetition,transformation,contradiction,consequence); u=clip(W_a a+W_l l+W_h h+W_r r+W_m m,u_min,u_max).

I(e)=−log P(e|context,history); J_A=λ_P P+λ_D D+λ_C C−λ_S S−λ_M M−λ_R R.

G_interact=(tracks,edges); label(edge)∈{answer,echo,console,shadow,contradict,anticipate,interrupt,reinforce,withdraw}.

M: P,D,C denote predictability, disconnection and gratuitous complexity; S,M,R denote meaningful surprise, motif recall and narrative relevance. These require declared measurement or judgment functions, not universal constants. Affect is an interpreted control signal, not sentience. Contradiction, silence and departures in phrase length, pickups, voicing, cadence or entry can serve the narrative.

## M06 • Historical and cultural atlas

θ_r=(region,era,community,occasion,tuning,meter,rhythm,form,grammar,ensemble,technique,production,affect,provenance).

Atlas = { (ID_r,θ_r,anchors_r,source_spans_r,rights_r,uncertainty_r) }; Flatten(community) is prohibited by the schema.

Tradition_r ≠ InstrumentClass_r; influence ≠ license; chronological order ≠ quality order.

M: the historical, global modal and recorded-genre inventories are finite reference registries, not universal coverage. Appendix A sections 6-8 preserve every named composer, performer, theorist, instrument, tradition, era and technique. Each source line is a complete retained record; expansion into evidence-backed individual records remains work, not an asserted completed research corpus.

## M07 • Composition and long-range structure

G_song=(sections,phrases,motifs,lyrics,affect arcs,harmony,tempo,meter,roles,transitions).

P(IR|C)=P(F,arc|C)P(motifs,L|F,C)P(Tune,H,Rhythm|motifs,C) × P(melody,prosody|L,H,Rhythm)P(ensemble|melody,arc)P(gestures|ensemble,performer)P(audio|gestures,acoustics,mix).

Section_k=(stakes_k,tension_k,lead_k,answer_k,rest_k,innovation_k); revision messages are bounded and versioned.

M: this is a proposed factorization, with conditional-independence assumptions rather than a demonstrated generative law. Development can establish, complicate, reframe, climax or anticlimax, then leave an afterglow. Long-range motif memory and intentional silence constrain filler. Backward revision is allowed without silently rewriting accepted inputs.

## M08 • Phonetic singing and fidelity

L → (language,dialect) → G2P → IPA → (syllables,stress) → breath → alignment → coarticulation → excitation → tract.

y=h_tract(phoneme,formants,register)*g_glottis($f_0$,tilt,breath)+frication; $f_0(s)=f_{base}(s)2^{(c(s)/1200)}$.

c(s)=vibrato(s)+scoop(s)+microbend(s); E_prosody=Σ_j(w_s E_stress+w_v E_vowel+w_c E_consonant+w_b E_breath+w_t E_timing).

Singer=(notes,phonemes,durations,expression,authorized identity,language); WER=(S+D+I)/N_ref.

M/E: the source-filter model is approximate; a neural singer need not literally implement it. Vowel sustain, consonant placement, register/passaggio, diction, accent, phonation and authorized identity are separate constraints. ASR is fallible on singing, and forced alignment alone cannot prove words were sung. Combine transcription, alignment, acoustic inspection and listening; log uncertainty per line.