Learning to Sight-Sing: The Mental Mechanics of Aural Imagery

Lee Humphries

Here is a mental strategy for translating musical notation into aural imagery. It solves sight-singing problems in both tonal and atonal melodies, yet it is easy to learn and use. Would-be sight-singers will find its logic laid out graphically in the paper's figures. The narrative explains the psychological principles that underlie them.

The strategy's power springs from its efficient structure, which compensates for the dual limitations of long-term memory and short-term memory. What are these limitations, and how does it offset them? Let's start at the very beginning.

An Overview of the Problem and Its Solution

Pitch. By virtue of the perceptual phenomenon of *octave equivalence*, tones whose frequencies are related by a factor of 2^x (x being a whole number) are heard as possessing the same quality. For example, the frequencies 440Hz and 1760Hz—*a′* and *a‴* respectively—are perceived to be alike, but of different register.

And by virtue of the perceptual phenomenon of *proximate equivalence*, tones whose frequencies are nearly equal are also heard as possessing the same quality. The frequencies 435Hz, 440Hz, and 445Hz are perceived to be different tunings within a single qualitative category—variants of *a′*.

Western music exploits these phenomena to reduce the audible frequency continuum to twelve distinct tonal qualities. All tones possessing the same quality are regarded as the same pitch. Symbolically, each pitch is represented by a unique set of enharmonically equivalent names: {*A*}, {*A#* or *Bb*}, {*B* or *Cb*}, {*B#* or *C*}, {*C#* or *Db*}, {*D*}, {*D#* or *Eb*}, {*E* or *Fb*}, {*E#* or *F*}, {*F#* or *Gb*}, {*G*}, {*G#* or *Ab*}. Musical context determines the name used.

Long-term memory. As independent tonal phenomena, the twelve pitches are perceptually challenging. We quickly forget which is which. Unless you are endowed with "absolute pitch" recognition—an unusual ability sometimes acquired with early musical training—associations of pitch and name won't stick in your long-term memory.

But all is not lost. Long-term memory is quite good at storing and differentiating *patterns* of pitches. Mention a familiar tune, and most people can sing it. What they recall is a series of pitch relationships, not a series of absolute pitches. This brings us to sight-singing's First Principle: *To process absolute pitches we must use pitch relationships as their proxies.*

Short-term memory. Our ability to hold information in consciousness is constrained by the *channel capacity* of short-term memory. Channel capacity is approximately 7±2 chunks of information. Ample evidence for this is presented in George A. Miller's classic article, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information" (*Psychological Review*, Vol. 63, No. 2, March 1956).

Because seven—more or less—is the maximum number of cognitive units we can simultaneously track and manipulate, any sight-singing framework that exceeds that number will be difficult to learn. This establishes sight-singing's Second Principle: *Collectively, the number of pitch relationships must not exceed 7±2 cognitive units.*

Given that there are twelve pitches, how is this possible?

Channel load. The demand—or *channel load*—that a collection of pitches places on short-term memory reflects their disconnectedness from one another. Pitches devoid of relationship take up more "space" in memory than pitches manifesting a relationship.

Fortunately, our perception of connectivity is malleable. With repetition, a series of initially independent pitches will coalesce into a coherent mental entity—a perceptually integrated configuration of relationships called a *gestalt*. A gestalt is a higher level cognitive unit. When individual pitches organize into a gestalt, their channel load is reduced. The reduction frees up space, enabling short-term memory to handle more pitches.

Gestalts can integrate into higher level gestalts. The second-order integration further reduces the channel load. This brings us to sight-singing's Third Principle: *For twelve pitches to occupy 7±2 cognitive units, their proxy relationships must be integrated into gestalts.*

Accordingly, we define the optimal sight-singing framework as *the "least-channel-load" assembly of gestalts capable of generating the aural images of all twelve pitches.*

The Logical Derivation of the Optimal Sight-Singing Framework

As we have said, the twelve pitches lack permanently memorable distinctions when heard as absolute values. To tell them apart, we must convert them into relational values. This is done in three steps.

Step 1. First, we contextualize the twelve pitches with respect to a *reference* pitch. For our reference, we select the pitch that is the melody's primary focus. If the melody is highly chromatic and lacks tonal focus, any convenient pitch will work.

Selecting one pitch as a reference puts every other pitch in a unique aural relationship to it. These contrasting relationships, called *functions*, distinguish the twelve pitches.

Function numbers. To discuss the different functions we must first name them. We'll use a "diatonic" numbering system based on the seven degrees of the major scale. (The major scale's ascending sequence of whole- and half-steps, *w* and *h* respectively, is *wwhwwwh*.) At the end of the paper, we'll present an alternative, "chromatic" system. Both are numeric "movable *Do*" systems.

In diatonic numbering, we name each function using one of the scale degrees (*1*, *2*, *3*, *4*, *5*, *6*, or *7*), either with or without one of two accidentals (# or *b*). We designate the reference pitch as function *1*. A number *by itself* refers to the corresponding degree of the major scale. For example, *6* is the function corresponding to the major scale's sixth degree.

Any number *with a preceding accidental* specifies a half-step deviation from that degree of the major scale—either higher (#) or lower (*b*). As a result, *#6* is the function a half-step higher than the major scale's sixth degree, and *b6* is the function a half-step lower. Since the sharp or flat in a function's number expresses only its half-step deviation from the major scale (and nothing else), it may refer to a pitch that has no such accidental. For example, when function *1* is *Bb*, *sharp-four* is *E-natural*; when function *1* is *A*, *flat-six* is *F-natural*.

A single function can be designated by various enharmonically-equivalent scale degrees. The enharmonic function names are: {*1*},{*#1* or *b2*},{*2*},{*#2* or *b3*},{*3* or *b4*},{*#3* or *4*},{*#4* or *b5*},{*5*},{*#5* or *b6*},{*6*},{*#6* or *b7*},{*7* or *b1*}. Musical context determines the name used.

A beginner's mind rebels when one thing goes by two names. So for now, we'll reject enharmonic aliases and give each function a single label. We prefer imprecise clarity to precise confusion. See Fig. 1.

1	b2	2	b3	3	4	#4	5	b6	6	b7	7

Fig. 1. The names of the twelve functions (diatonic system)

Step 2. The translation from an absolute pitch to its relative function moves us closer to our goal. But the twelve functions still exceed the 7±2 channel capacity of short-term memory. Further translation is needed.

Let us re-contextualize the twelve functions. Divide the twelve functions into two related, but mutually exclusive, groups. Call one group *landmark functions*, and call the other group *non-landmark functions*. Think of each landmark function as a point. And think of each non-landmark function as an interval of distance above or below such a point.

Because there are multiple landmark points, the number of intervals needed to locate the non-landmark functions is fewer than the number of non-landmark functions themselves. Exploiting this reduction, we recast the non-landmark functions as intervals. The recasting leaves us with two classes of aural elements: *landmarks* and *intervals from landmarks*.

We want to do the most with the least. Therefore, we optimize the membership of each class. We limit the landmark class to *functions that (1) collectively are consonant with each other and (2) individually reinforce the reference pitch*. And we limit the interval class to *the fewest and smallest landmark-originating intervals needed to produce all the non-landmark functions*.

It is easy to demonstrate that the optimal landmark class has only three elements and that the optimal interval class has only four elements. In total, we have but seven aural elements to manipulate in consciousness. This approximates the channel capacity of short-term memory.

Landmark elements. The three-element landmark class has two possible forms—either of which can support our sight-singing framework. One form contains functions *1*, *3*, and *5*. The other form contains functions *1*, *b3*, and *5*. In both forms, the three landmarks are mutually consonant, and function *1* is reinforced as reference by its root relationship to the other two functions. Thus reinforced, function *1* acquires the aural role of *tonic*, and the other functions take on—with respect to it—the aural roles of *major-mediant* (function *3*), *minor-mediant* (function *b3*), and *dominant* (function *5*).

Interval elements. The four-element interval class has only one form. It contains the descending half-step, the descending whole-step, the ascending half-step, and the ascending whole-step.

Step 3. When cognitively integrated, the three elements of either landmark class become a gestalt. The two forms produce, respectively, the *major triad* and *minor triad*. Likewise, when cognitively integrated, the four elements of the interval class become a gestalt, *the neighbor set*.

Gestalt-element symbols. Shortly, we will depict the sight-singing framework graphically. But before we can do that we must assign the gestalt elements unique symbols. We symbolize the elements of the triadic gestalts like this. The tonic landmark is *T*, the major-mediant landmark is *M*, the minor-mediant landmark is *m*, and the dominant landmark is *D*. Accordingly, the major-triad gestalt is {*T, M, D*}, and the minor-triad gestalt is {*T, m, D*}.

The elements of the neighbor set we symbolize like this. The descending whole-step interval is *–w*, the descending half-step interval is *–h*, the ascending half-step interval is *+h*, and the ascending whole-step interval is *+w*. So, the neighbor-set gestalt is {*–w, –h, +h, +w*}.

It is important to distinguish function *1* in its two structural roles: as the complete framework's fundamental reference, symbolized *R*, and as the triad gestalt's tonic landmark, symbolized previously as *T*.

***

Let's summarize. What have we done, and where has it brought us? We have taken twelve pitches, reincarnated them as twelve functions, then transmuted them into three landmarks and four intervals. Finally, we have organized the landmarks into a triad gestalt and the intervals into a neighbor gestalt. As a result, the twelve functions are ready to take on discernible aural roles within a simple *syntactical framework*.

The Syntactical Framework

The syntactical framework is a *modular assembly* of aural gestalts—a triad gestalt (either major or minor) interfacing with three separate but identical neighbor gestalts.

Since the three neighbor gestalts are all alike, the assembly employs only two different components, rendering its channel load considerably less than the 7±2 capacity of short-term memory. Consciousness can juggle all this aural imagery at once, manipulating it as needed to solve any sight-singing problem that arises.

Picturing the Syntactical Framework

Origins. Our syntactical framework requires one more concept—the idea of an *origin*. We define an *origin* as any pitch: (1) on which a tonal gestalt is built and (2) with respect to which its other elements acquire aural meaning. We place the caret symbol (^) above a pitch to show that it is an origin. In the following figures (Figs. 2 through 5a), we base the syntactical framework on a major triad gestalt. Figures based on the minor triad gestalt will follow.

The origin of the triad gestalt is the reference pitch. See Fig. 2.

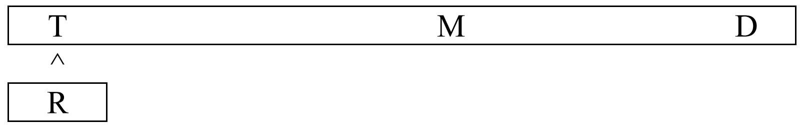


Fig. 2. The triad gestalt {T-M-D} originating on the reference R

The origin of the neighbor gestalt is any of the triad gestalt's three element. Thus, the neighbor gestalt has *three iterations*. See Fig. 3.

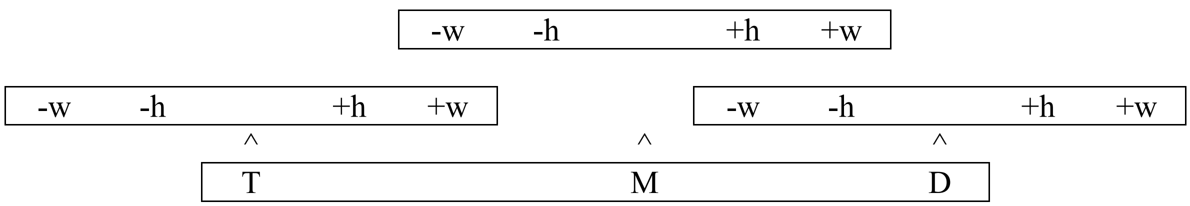


Fig. 3. The neighbor gestalt {-w, -h, +h, +w}
originating on each element of the triad gestalt {T-M-D}

It is the three points of attachment between the two gestalts that enable them to encompass all twelve functions. See Fig. 4. (Arranged in syntactical order, functions *b7* and *7* precede function *1* because they are its lower neighbors.)

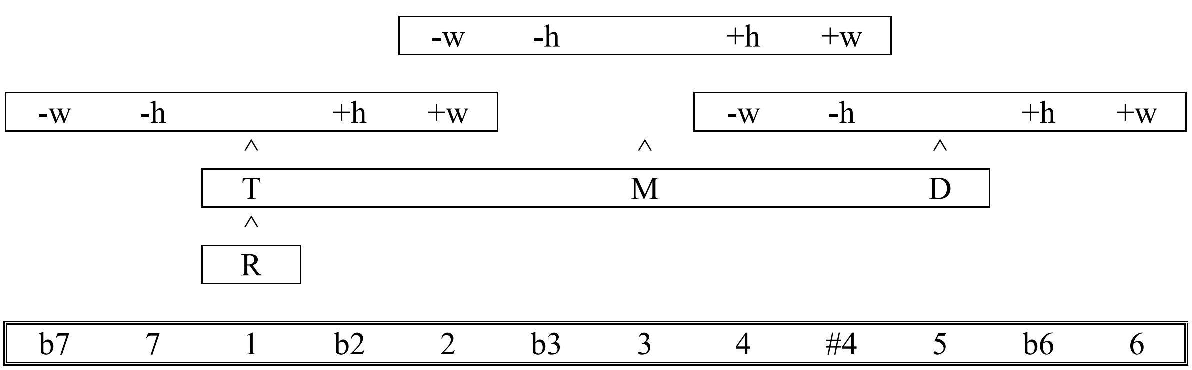


Fig. 4. The twelve functions encompassed by
the assembly of triad and neighbor gestalts

Addresses. Within the syntactical framework, each function has an *address*. The address gives the function's aural landmark (*T*, *M*, or *D*) and, when appropriate, its interval therefrom (*-w*, *-h*, *+h*, or *+w*). A few functions have two addresses because they can be approached in two ways. For example, if the triad gestalt is major, function *2* is either *T+w* or *M-w*. Figs. 5a and b show the complete syntactical framework (with functions on the bottom row and their corresponding addresses on the top). In Fig. 5a the triad is major; in Fig. 5b, minor.

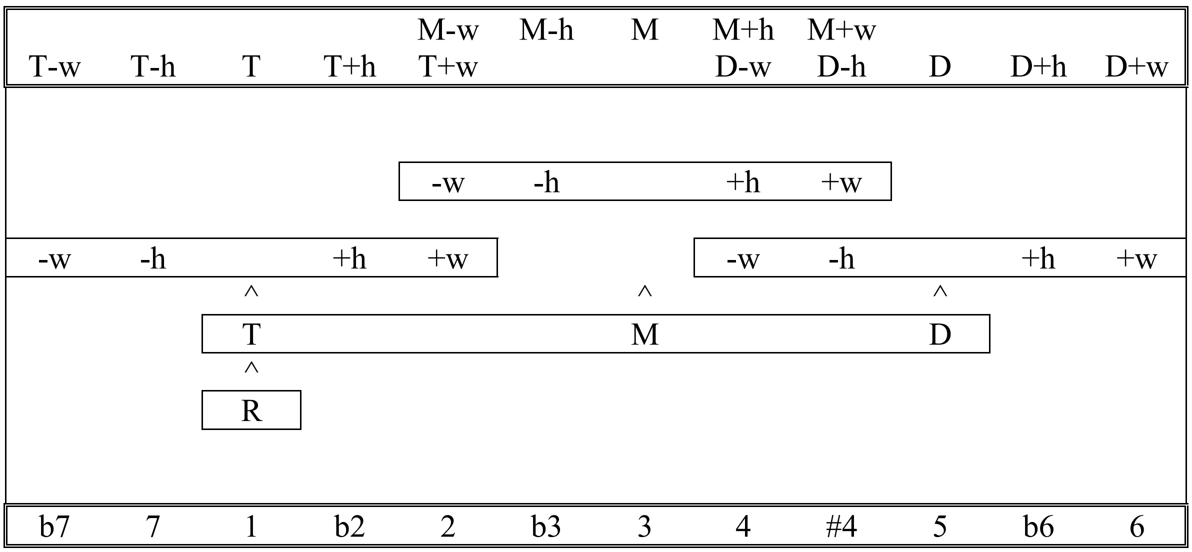


Fig. 5a. Function addresses within a major-triad syntactical framework

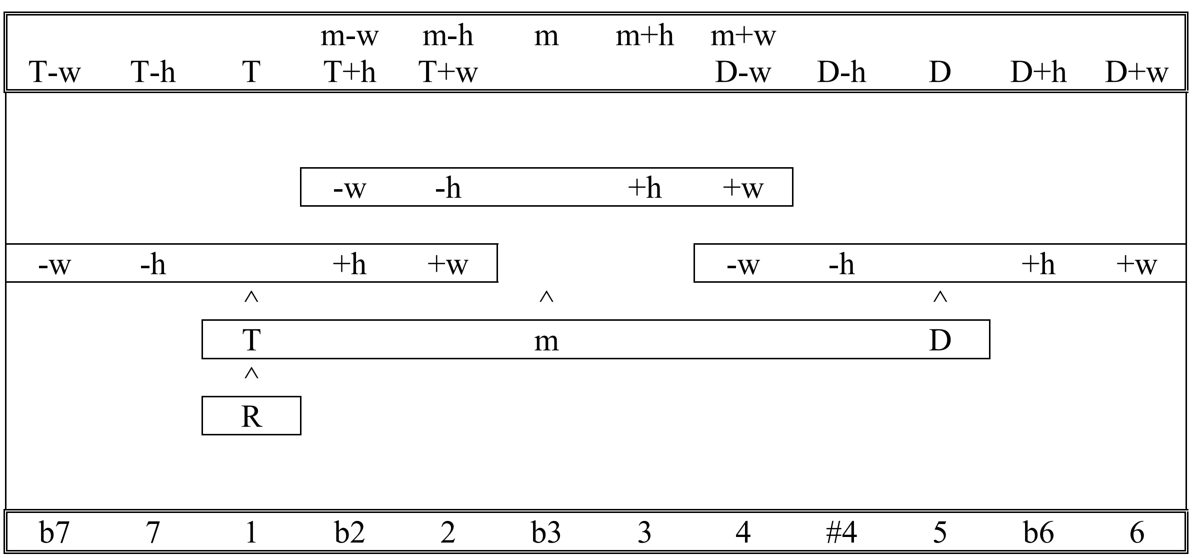


Fig. 5b. Function addresses within a minor-triad syntactical framework

Aural paths. A function's sound can be computed mentally by moving through a sequence of aural images. As seen in Figs. 5a and b, the path to each function begins in the same way. It starts at the reference pitch and proceeds through the triad gestalt to the appropriate landmark. The critical last leg is summarized by the function's address, which specifies the landmark and appropriate interval therefrom (if any). (In these figures, the triad is projected upward from the reference pitch. Shortly, we'll consider its projection downward. The computational principle remains the same.)

States of consciousness. Because each step along the path calls into consciousness a different aural image, a function's computation requires a specific sequence of mental states. This reveals sight-singing's Fourth Principle (best stated in the negative): *The inability to sing a function is either (1) a failure to know its derivational path or (2) a failure to recall a correct aural image somewhere along that path.* Once we recognize this, it's easy to locate the source of a sight-singing error.

Acquiring the Syntactical Framework's Aural Images

Singing is remarkable in its capacity to simultaneously generate two disparate streams of information: one tonal, the other linguistic. When we sing a function while voicing its number (e.g., "flat-seven"), we are mentally merging an aural relationship with its name. If we do this repeatedly, the two will meld in the mind, fusing into one blended object. After that, the appearance of one component will trigger a recall of the other. The complementary triggers advance our musical hearing.

The task, therefore, is two-fold: (1) to merge a set of aural functions with their numbers, and (2) to group those blended objects into a gestalt. In general terms, this is how you go about it:

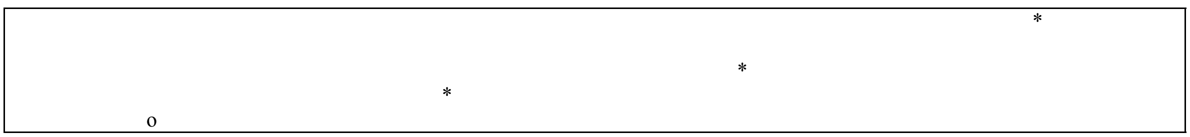
(1) *Present* to yourself—at the piano, say—a set of pitches that embody specific aural functions and are paired with the functions' numbers. (What matters are the functions themselves, not the particular pitches that convey them. Thus, the landmark functions of the major triad might be rendered as {C, E, G} or {D, F#, A}or some other congruent set of pitches.)
(2) *Imitate* the presentation vocally, singing each function's number.
(3) *Repeat* the imitation.

This simple chain of behaviors lays the foundation for the gestalt's later recall. *You cannot imitate a presentation without remembering it, and you cannot repeat the imitation without reinforcing the memory.*

Our first concerns are the triad gestalts and the neighbor gestalt. We approach them in the following order.

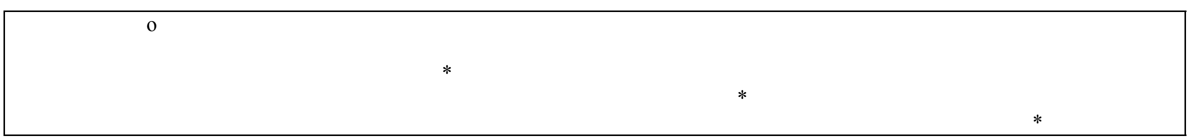
The triad gestalt. The lowest gestalt in the syntactical framework is the triad of landmark functions—in either major or minor form. Therefore, the triad gestalts must be learned first. Initially, we'll present each gestalt's elements with respect to its origin pitch. But our ultimate aim is to hear each element as a function of the reference pitch. For that reason, we sing the elements on their function numbers.

Below are depictions of the major and minor versions of the triad gestalt, with the landmarks labeled functionally and configured in both ascending and descending order. In ascending order, the gestalt's origin (o) is on the bottom (Fig. 6a); in descending order, it is on the top (Fig. 6b). Knowing both configurations enables us to project the triad gestalt upward or downward into an adjacent octave (as required by the range of a given melody). Play a gestalt, then sing it on numbers using any convenient triad; it is only the aural relationships that interest us.



1	3	5	1
1	(b)3	5	1

Fig. 6a. The ascending triad gestalt with its elements labeled functionally.



1	5	3	1
1	5	(b)3	1

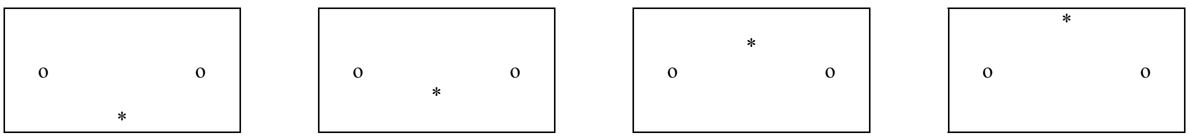
Fig. 6b. The descending triad gestalt with its elements labeled functionally.

The neighbor gestalt. Within the syntactical framework, the neighbor gestalt resides one level above the triad gestalt—because its origin is a member of the triad gestalt. Until the aural image of the triad gestalt is in place, aural construction of a neighbor gestalt is impossible.

We secure the neighbor gestalt in each of its syntactical positions, concentrating on one landmark origin at a time. In keeping with our ultimate aim to hear each gestalt element as a function of the reference pitch, we sing the gestalt elements {*–w, –h, +h, +w*} as, respectively, {*b7, 7, b2, 2*} when originating on *1*; as {*2, b3, 4, #4*} when on *3*; as {*b2, 2, 3, 4*} when on *b3*; and as {*4, #4, b6, 6*} when on *5*.

We want to accomplish two things: (1) associate each neighbor with its landmark origin and (2) differentiate one neighbor from another. To associate a neighbor with its landmark, we sing the *neighbor sequence* "origin-neighbor-origin"—e.g., "1-b7-1." To differentiate the four neighbors, we sing in close succession the four neighbor patterns of common origin—e.g., "1-b7-1," "1-7-1," "1-b2-1," "1-2-1." Differentiating these similar functions integrates them into a relational gestalt.

Fig. 7 depicts the four "origin-neighbor-origin" shapes and their functions relative to each triad landmark. Play a gestalt then sing it on numbers using any convenient triad to establish the landmarks.



5	4	5
3	2	3
b3	b2	b3
1	b7	1

5	#4	5
3	b3	3
b3	2	b3
1	7	1

5	b6	5
3	4	3
b3	3	b3
1	b2	1

5	6	5
3	#4	3
b3	4	b3
1	2	1

Fig. 7. "Origin-neighbor-origin" shapes around each triad landmar

Practicing aural paths. The next figures, Figs. 8a and b, show the sequence of aural images that will carry you to each target function. In Fig. 8a the triad gestalt is major; in Fig. 8b, it is minor. Since the target function can lie either above or below the reference pitch, we learn to project the triad gestalt in both directions.

Our objective is to internalize each function's computation, singing only the target function. We proceed in stages. First, we play then sing the complete path shown in Figs. 8 (a or b). Next, we sing the figure's parenthesized functions softly and the target function in full voice. Finally, we think the sounds of the parenthesized functions, singing only the target function. The parenthesized functions are *aural scaffolding* to put us within reach of the target function.

Aural Paths to Target Functions via the Major Triad Gestalt				
Target Function	lies above the reference pitch (ascending path)		lies below the reference pitch (descending path)	
1	(1-3-5) 1			
7	(1-3-5-1) 7		(1) 7	
b7	(1-3-5-1) b7		(1) b7	
6	(1-3-5) 6		(1-5) 6	
b6	(1-3-5) b6		(1-5) b6	
5	(1-3) 5		(1) 5	
#4	(1-3-5) #4	(1-3) #4	(1-5) #4	(1-5-3) #4
4	(1-3-5) 4	(1-3) 4	(1-5) 4	(1-5-3) 4
3	(1) 3		(1-5) 3	
b3	(1-3) b3		(1-5-3) b3	
2	(1) 2	(1-3) 2	(1-5-3) 2	(1-5-3-1) 2
b2	(1) b2		(1-5-3-1) b2	
1			(1-5-3) 1	

Fig. 8a. Aural scaffolding: The path to target functions via the major triad gestalt

Aural Paths to Target Functions via the Minor Triad Gestalt				
Target Function	lies above the reference pitch (ascending path)		lies below the reference pitch (descending path)	
1	(1-b3-5) 1			
7	(1-b3-5-1) 7		(1) 7	
b7	(1-b3-5-1) b7		(1) b7	
6	(1-b3-5) 6		(1-5) 6	
b6	(1-b3-5) b6		(1-5) b6	
5	(1-b3) 5		(1) 5	
#4	(1-b3-5) #4		(1-5) #4	
4	(1-b3-5) 4	(1-b3) 4	(1-5) 4	(1-5-b3) 4
3	(1-b3) 3		(1-5-b3) 3	
b3	(1) b3		(1-5) b3	
2	(1-b3) 2	(1) 2	(1-5-b3) 2	(1-5-b3-1) 2
b2	(1-b3) b2	(1) b2	(1-5-b3) b2	(1-5-b3-1) b2
1			(1-5-b3) 1	

Fig. 8b. Aural scaffolding: The path to target functions via the minor triad gestalt

A demonstration. The table and the musical notation in Fig. 9 show you how to implement aural paths to sing an atonal twelve-pitch series. Arbitrarily, we choose the pitch *F* for our reference. We designate *f′* (the *F* above middle *C*) as our framework's origin and mentally erect on it a major triad.

The table's top row gives the pitch series. Its middle row gives the pitches' functions with respect to *F*. And its bottom row gives the ascending (^) or descending (v) aural path through which the pitches are computed. To insure that each step is clearly imaged before proceeding to the next, we move along the path slowly.

Below the table the same information is displayed in musical notation. The twelve-pitch series is depicted in whole notes. The aural path to a pitch is depicted by the grace notes that precede it.

Pitch	a	g′	d″	$e^{b\prime}$	c″	$g^{b\prime}$	$b^{b\prime}$	b	$d^{b\prime}$	$a^{b\prime}$	f′	e′
Function	3	2	6	b7	5	b2	4	#4	b6	b3	1	7
Path	v(1-5) 3	^(1) 2	^(1-3-5) 6	v(1) b7	^(1-3) 5	^(1) b2	^(1-3) 4	v(1-5) #4	v(1-5) b6	^(1-3) b3	1	v(1) 7

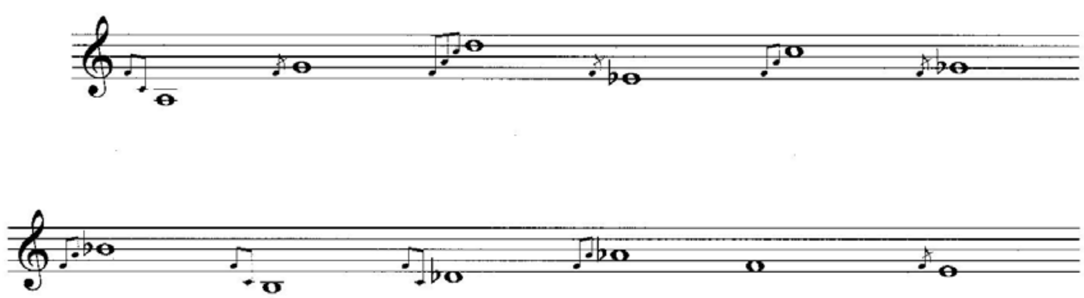


Fig. 9. An atonal pitch series and its aural computation

Manipulating the Syntactical Framework's Aural Images

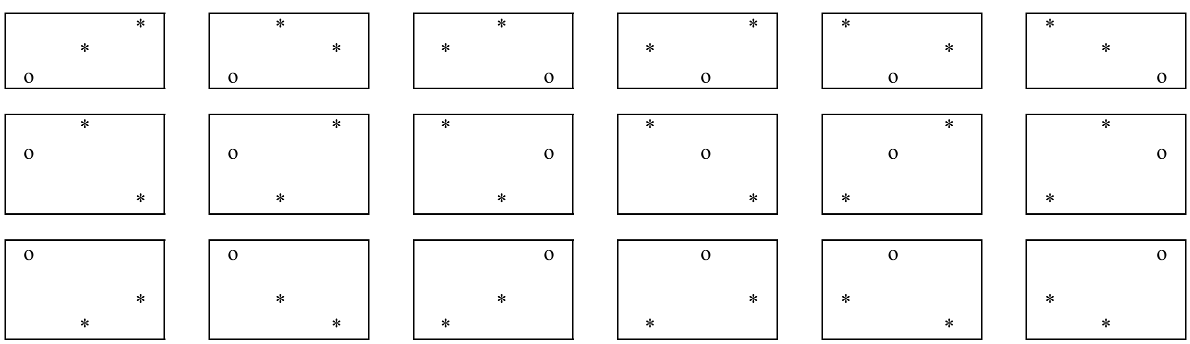
Mastering the materials in Figs. 6, 7, and 8 will give you a coherent computational framework. Mastering their elemental transformations and combinations (which follow) will enhance its flexibility.

Learning to manipulate the basic gestalts. As treated above, the gestalts are *ordered*—their elements occur in a fixed sequence. We now consider their rearrangements, called *permutations*.

Triad permutations. In Fig. 10, each row contains six patterns. These are re-orderings of a single set of triad elements.
Practice the complete set of patterns as a major triad, then as a minor triad, proceeding like this. Play the first pattern and sing it on numbers until its three pitches are comfortably lodged in your short-term memory. Then, continuing in the same row, sing the five patterns that follow. In doing so, you will be mentally permuting recollected aural images. This will enhance your ability to perceive the triad gestalt as an *unordered* set of aural relationships—a generalized cognitive entity that subsumes all particular arrangements of its elements.

Each column of Fig. 10 contains three patterns. Each pattern presents the triad elements in the same order but in different registers. In the first pattern, the gestalt's origin is on the bottom. In the second pattern, it is in the middle. In the third, on the top.

Practice the complete set of patterns as a major triad, then as a minor triad, proceeding like this. Play the first pattern and sing it on numbers until its three tones are comfortably lodged in short-term memory; then sing patterns two and three, which appear beneath it. Repeat the procedure, traveling upward from the bottom pattern. This task requires you to mentally substitute one aural image for another, replacing a pitch with its octave equivalent. These octave substitutions will enhance your ability to perceive the triad gestalt as a *non-localized* set of aural relationships, mentally projectable into any register.



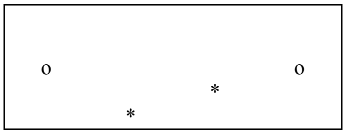
1 3 5	1 5 3	3 5 1	3 1 5	5 1 3	5 3 1
1 (b)3 5	1 5 (b)3	(b)3 5 1	(b)3 1 5	5 1 (b)3	5 (b)3 1

Fig. 10. The triad gestalt: The three triad landmarks,
permuted (rows) and displaced (columns)

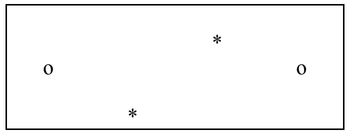
Familiarity with these permutations and displacements will simplify a target function's aural path. No longer must you move sequentially through the triad elements to reach the appropriate landmark. *The imagery is reduced to the function's syntactical address: landmark plus interval (if any).*

Neighbor permutations. The patterns in Fig. 11 (a, b, c, and d) take the form "origin-neighbor$_i$-neighbor$_j$-origin"—e.g., "1-b7-7-1." Here, we juxtapose two of a landmark's four neighbors, presenting them in all possible permutations. In singing these patterns, we develop the capacity to digress from a landmark while preserving it as an aural anchor. Musical hearing improves as we learn to sing extended series of non-triad functions without losing our orientation to the aural framework's reference pitch.

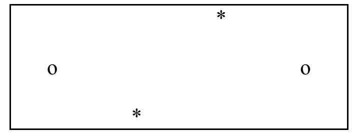
a.



5	4	#4	5
3	2	b3	3
b3	b2	2	b3
1	b7	7	1

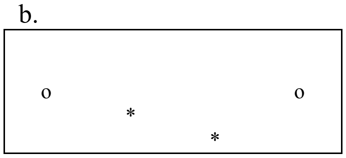


5	4	b6	5
3	2	4	3
b3	b2	3	b3
1	b7	b2	1

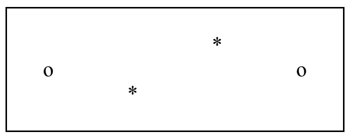


5	4	6	5
3	2	#4	3
b3	b2	4	b3
1	b7	2	1

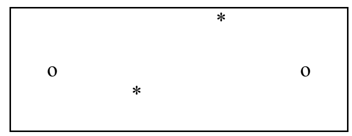
b.



5	#4	4	5
3	b3	2	3
b3	2	b2	b3
1	7	b7	1

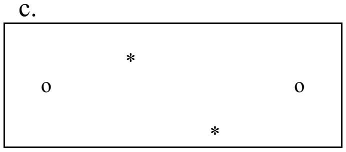


5	#4	b6	5
3	b3	4	3
b3	2	3	b3
1	7	b2	1

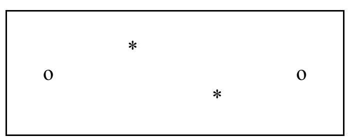


5	#4	6	5
3	b3	#4	3
b3	2	4	b3
1	7	2	1

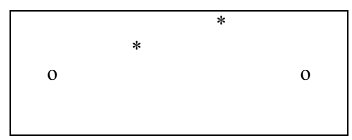
c.



5	b6	4	5
3	4	2	3
b3	3	b2	b3
1	b2	b7	1

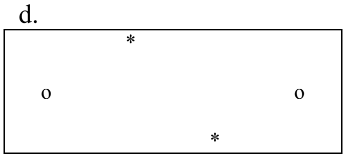


5	b6	#4	5
3	4	b3	3
b3	3	2	b3
1	b2	7	1

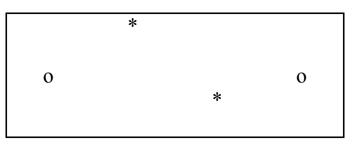


5	b6	6	5
3	4	#4	3
b3	3	4	b3
1	b2	2	1

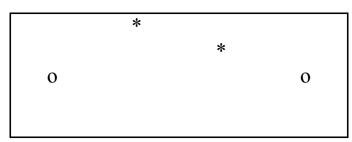
d.



5	6	4	5
3	#4	2	3
b3	4	b2	b3
1	2	b7	1



5	6	#4	5
3	#4	b3	3
b3	4	2	b3
1	2	7	1



5	6	b6	5
3	#4	4	3
b3	4	3	b3
1	2	b2	1

Fig. 11. "Origin-neighbor$_i$-neighbor$_j$-origin" shapes around each triad landmark

A Pedagogical Approach to Aural Imagery

Passive acquisition of the framework. The above discussion explains how to acquire the syntactical framework on your own. But, in truth, it is more easily acquired in an interactive group setting that is metrically structured as a *call-and-response*.

To a steady beat, the instructor rhythmically sings a gestalt on numbers, and the learners (in cadence) imitate, then repeat again, what they have heard. Maintaining the beat, the instructor sings another gestalt and the learners again respond, etc. Immersed in this metrical environment, the learners can rapidly experience the triad gestalts in all their permutations, followed by each landmark's neighbor gestalt in all its permutations.

The advantages of the call-and-response structure are significant. First, the requirement to consciously merge a function with its number is obviated; the mental fusion is established automatically as learners imitate what they have heard. Second, learners become physiologically and psychologically *entrained* to the beat and to the metric points where tonal-linguistic information occurs. As an information point draws near, expectation heightens and attention focuses; as information is transmitted, expectation is fulfilled and closure occurs. The recurring antecedent-consequent form— "expectation-fulfillment-expectation-fulfillment"— becomes a hypnotic *"carrier" wave*. This allows its embedded information to bypass the rational mind and register more easily in the unconscious.

In this setting, learners with a modicum of musical experience can acquire the basic aural framework very quickly—often in one session.

Translating pitch into function. Surprisingly, the aural part of the process is the easiest to learn. More troublesome is the linguistic part: the mental translation of pitch name into function name. Since any of the twelve pitches can serve as the reference, each pitch can assume twelve different functions.

It takes practice to know instantly the function of any pitch with respect to any other. Yet even this becomes easier when you picture each pitch as an address within a syntactical framework. This parsing reduces your prerequisite knowledge to just a few rudimentary things: the spelling of major and minor triads, and the spelling of half- and whole-step intervals.

Aural Evolution

The Evolution of the Syntactical Framework

With experience, the network of connections within your aural framework will grow richer. Learners who practice Figs. 10 and 11 will streamline their ability to navigate within it. Further enrichment will occur as recurring sequences of functions coalesce into new gestalts.

Some of these gestalts add a higher level of organization to the framework. If a new gestalt anchors itself to a neighbor element—instead of a triad element—that neighbor becomes a *secondary landmark*, an origin of a higher order. Take, for example, the function sequence *7-2-4-b6*. Heard often enough, these functions will merge into a common gestalt, the *diminished-seventh chord*. Thus integrated, they can be processed en masse—as a coherent aural entity originating on function *7*. This is depicted in Fig. 12. The arbitrary signs *@*, *&*, *%*, and *$* represent the four elements of the diminished-seventh chord gestalt.

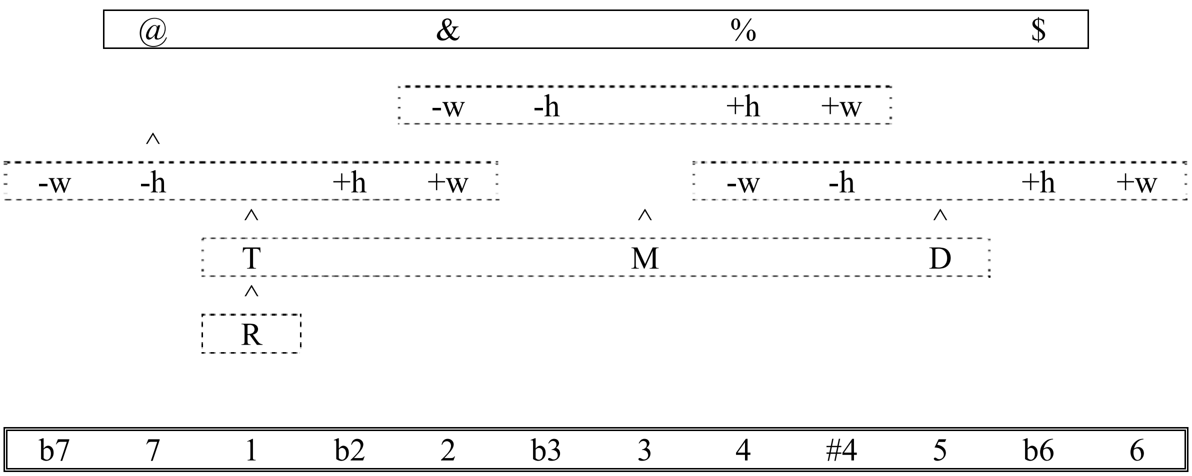


Fig. 12. A new gestalt (diminished-seventh chord)
originating on a secondary landmark (function *7*)

Fueled by the successful imaging and integration of ever more function patterns, our mental framework grows into a complex network of interconnected gestalts. As it does, we process aural information with increasing efficiency.

The Transition from "Movable *Do*" to "Fixed *Do*."
The final stage in our aural evolution is the transition to a system of *stationary* relationships. In making the transition, we progress from a numeric "movable *do*" system (where the reference pitch is *variable*) to a syllabic "fixed *do*" system (where pitches are once again regarded as *absolute*, their fixed aural relationships superimposed on the functions they represent).

Pitch syllables. Within the stationary system, we forego enharmonic pitch names (e.g., *C#* vs. *Db*). They slow aural processing. Instead, we designate the twelve pitches, whatever their spelling, with twelve fixed *syllables*. These are, in ascending chromatic order from the pitch *A*: *La*, *Be*, *Ti*, *Do*, *Na*, *Re*, *Go*, *Mi*, *Fa*, *Ke*, *So*, and *Vi*. Thus, whether a pitch is notated as B^x, *C#*, or *Db*, its syllable is always *Na*.

The syllables' vowel sounds conform to Latin pronunciation: *a* = ah, *e* = ay, *i* = ee, *o* = oh. (The black note names—*Na*, *Go*, *Ke*, *Vi*, and *Be*—were devised in the 1960s by the great musician and educator Allen Winold.) Fig. 13 shows the correspondence between pitch names and syllables.

Pitch	A	A#/Bb	B/Cb	C	C#/Db	D	D#/Eb	E/Fb	E#/F	F#/Gb	G	G#/Ab
Syll.	La	Be	Ti	Do	Na	Re	Go	Mi	Fa	Ke	So	Vi

Fig. 13. Enharmonic pitch names and their corresponding syllables

Sound-syllable mapping. In learning to sight-sing, we form a chain of paired associations, mapping pitch names onto function names and function names onto aural images. When the sequence is established, our recognition of a notated pitch's function triggers its aural image.

Having reached this level of development, we are prepared to form one more association—mapping aural images onto syllables. This fusion of sound and syllable creates another kind of blended object.

The mapping is most easily acquired by singing a melody on function numbers until they flow effortlessly; then singing the same melody on chromatic syllables, aiming for the same ease.

With sufficient repetition, patterns of sound-syllable objects become aural gestalts in their own right. Those gestalts permit us to generalize the aural forms that underlie them—to hear the sounds of unusual function patterns as transpositions of familiar function patterns. This is how it works.

Intervallic form. Each function pattern has an *intervallic form*. For example, the function pattern *1-3-5* has the intervallic form *major third + minor third*. Moreover, different function patterns can share the same intervallic form. For example, *#4-#6-#1* has the same form as *1-3-5*. Such patterns are transpositions of one another.

A syllable pattern also has an intervallic form. And each syllable pattern—along with the intervallic form it embodies—is applicable to twelve different function patterns. (The application is determined by the reference pitch, as shown below.)

Now, the intervallic form of the syllable pattern *Do-Mi-So* is identical to that of the function patterns *1-3-5* and *#4-#6-#1* (as well as ten others). Thus, with a reference of *C*, the syllable pattern *Do-Mi-So* will occur as *1-3-5*. And with a reference of *Gb*, it will occur as *#4-#6-#1*.

Rare vs. familiar. Some function patterns (like *1-3-5*) occur frequently, while others (like *#4-#6-#1*) occur rarely. Rare patterns are harder to image, but syllables can smooth the way.

That is possible when the syllables for a *rare* function pattern have been previously associated with a *familiar* function pattern. Then the aural image of the familiar intervallic form can transfer to the rare pattern via their shared syllables.

Therefore, if the function pattern *1-3-5* (reference *C*) is securely mapped onto the syllables *Do-Mi-So*, it becomes possible to re-map the syllables *Do-Mi-So* (and the intervallic form they embody) onto its transposition, *#4-#6-#1* (reference *Gb*). See Fig. 14.

Intervallic form	Major third		Minor third
Syllables	Do	Mi	So
Functions in C	1	3	5
Functions in Gb	#4	#6	#1

Fig. 14. Two congruent function patterns and their common syllable pattern

In general, when two function patterns are intervallically congruent and share the same syllable pattern, the syllables can superimpose the aural image of the familiar functions onto the rare functions.

Appendix: Naming Functions Chromatically

If we number the functions chromatically instead of diatonically, we avoid a double-edged sword: (1) the problem of calling one aural function by two names (e.g., *#1* or *b2*) versus (2) the problem of mislabeling an aural function's musical role (e.g., calling the function *b2* when it resolves as *#1*).

In the chromatic system, we designate the twelve functions using twelve sequential numbers—*0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11*. The reference pitch is function *0*. The other pitches are numbered as they would occur in an ascending chromatic scale built on the reference.

Fig. 15 shows the correspondence between function numbers in the diatonic and chromatic systems.

Diatonic	1	#1/b2	2	#2/b3	3	4	#4/b5	5	#5/b6	6	#6/b7	7
Chromatic	0	1	2	3	4	5	6	7	8	9	10	11

Fig. 15. Diatonic function names and their corresponding chromatic names

In the chromatic system, the major triad gestalt is {*0, 4, 7*} and the minor triad gestalt is {*0, 3, 7*}. For vocal facility, we sing *0* as "oh," *7* as "sev," and *11* as "lev."

***

We conclude by re-presenting important earlier figures with their functions named chromatically. No comment is necessary.

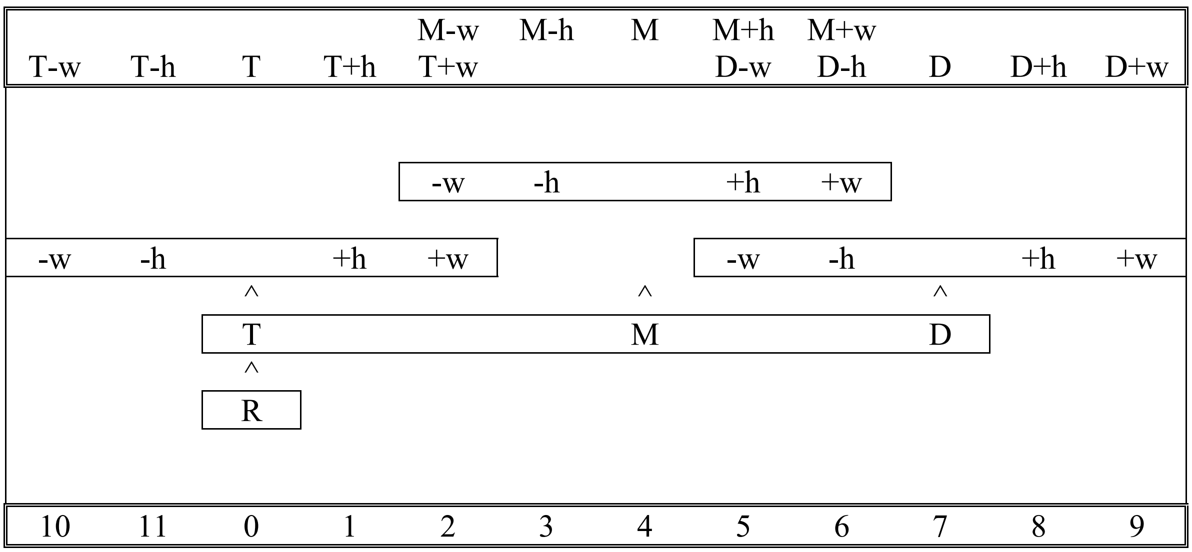


Fig. 16a. Function addresses within a major-triad syntactical framework

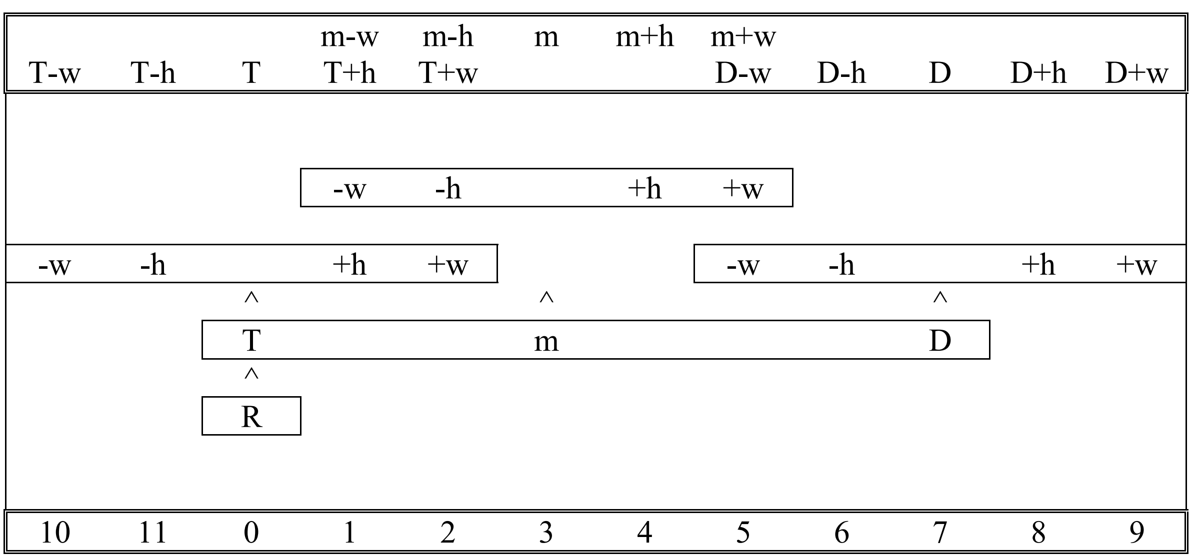
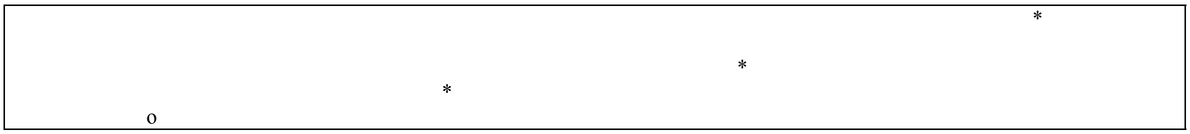
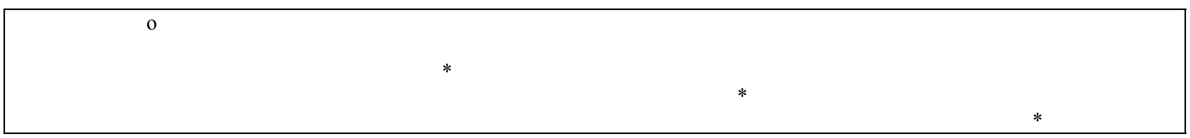


Fig. 16b. Function addresses within a minor-triad syntactical framework



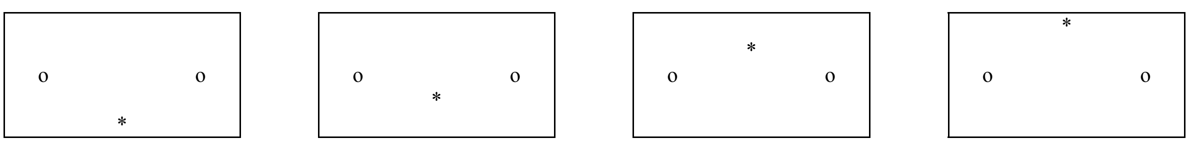
0	4	7	0
0	3	7	0

Fig. 17a. The ascending triad gestalt with its elements labeled functionally.



0	7	4	0
0	7	3	0

Fig. 17b. The descending triad gestalt with its elements labeled functionally.



7	5	7
4	2	4
3	1	3
0	10	0

7	6	7
4	3	4
3	2	3
0	11	0

7	8	7
4	5	4
3	4	3
0	1	0

7	9	7
4	6	4
3	5	3
0	2	0

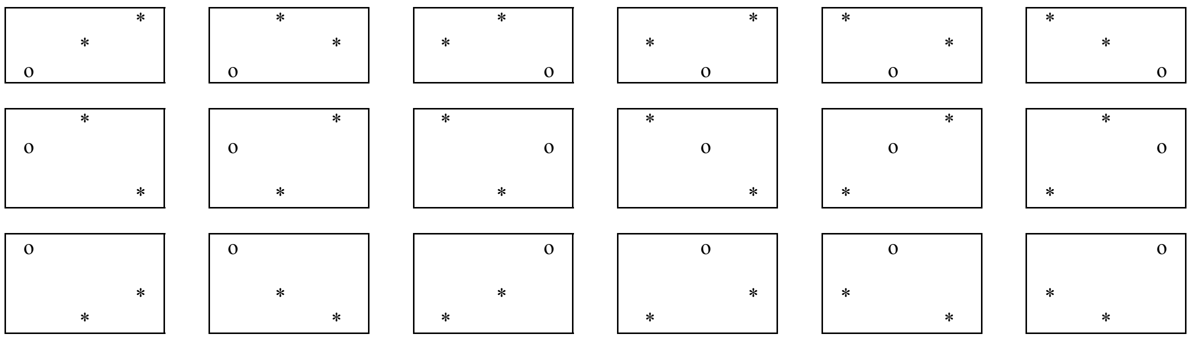
Fig. 18. “Origin-neighbor-origin” shapes around each triad landmark

Aural Path to Target Functions via the Major Triad Gestalt				
Target Function	lies above the reference pitch (ascending path)		lies below the reference pitch (descending path)	
0	(0-4-7) 0			
11	(0-4-7-0) 11		(0) 11	
10	(0-4-7-0) 10		(0) 10	
9	(0-4-7) 9		(0-7) 9	
8	(0-4-7) 8		(0-7) 8	
7	(0-4) 7		(0) 7	
6	(0-4-7) 6	(0-4) 6	(0-7) 6	(0-7-4) 6
5	(0-4-7) 5	(0-4) 5	(0-7) 5	(0-7-4) 5
4	(0) 4		(0-7) 4	
3	(0-4) 3		(0-7-4) 3	
2	(0) 2	(0-4) 2	(0-7-4) 2	(0-7-4-0) 2
1	(0) 1		(0-7-4-0) 1	
0			(0-7-4) 0	

Fig. 19a. Aural scaffolding: The path to target functions via the major triad gestalt

Aural Path to Target Functions via the Minor Triad Gestalt				
Target Function	lies above the reference pitch (ascending path)		lies below the reference pitch (descending path)	
0	(0-3-7) 0			
11	(0-3-7-0) 11		(0) 11	
10	(0-3-7-0) 10		(0) 10	
9	(0-3-7) 9		(0-7) 9	
8	(0-3-7) 8		(0-7) 8	
7	(0-3) 7		(0) 7	
6	(0-3-7) 6		(0-7) 6	
5	(0-3-7) 5	(0-3) 5	(0-7) 5	(0-7-3) 5
4	(0-3) 4		(0-7-3) 4	
3	(0) 3		(0-7) 3	
2	(0-3) 2	(0) 2	(0-7-3) 2	(0-7-3-0) 2
1	(0-3) 1	(0) 1	(0-7-3) 1	(0-7-3-0) 1
0			(0-7-3) 0	

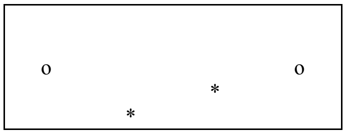
Fig. 19b. Aural scaffolding: The path to target functions via the minor triad gestalt



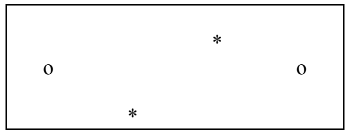
0 4 7	0 7 4	4 7 0	4 0 7	7 0 4	7 4 0
0 3 7	0 7 (3	3 7 0	3 0 7	7 0 3	7 3 0

Fig. 20. The triad gestalt: The three triad landmarks,
permuted (rows) and displaced (columns)

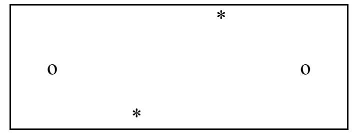
a.



7	5	6	7
4	2	3	4
3	1	2	3
0	10	11	0

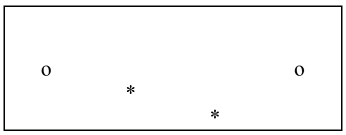


7	5	8	7
4	2	5	4
3	1	4	3
0	10	1	0

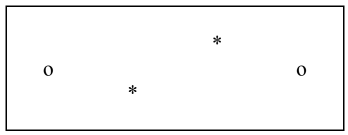


7	5	9	7
4	2	6	4
3	1	5	3
0	10	2	0

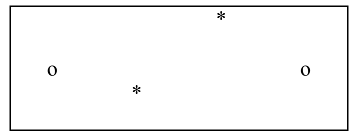
b.



7	6	5	7
4	3	2	4
3	2	1	3
0	11	10	0

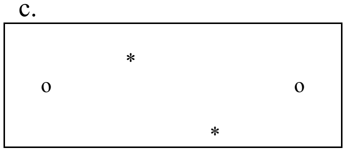


7	6	8	7
4	3	5	4
3	2	4	3
0	11	1	0

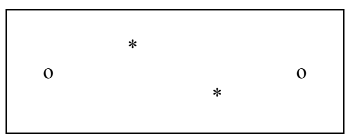


7	6	9	7
4	3	6	4
3	2	5	3
0	11	2	0

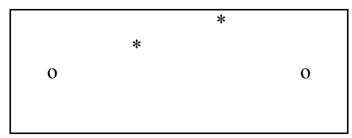
c.



7	8	5	7
4	5	2	4
3	4	1	3
0	1	10	0

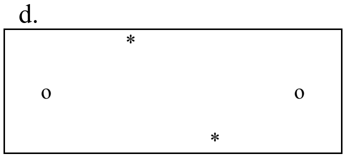


7	8	6	7
4	5	3	4
3	4	2	3
0	1	11	0

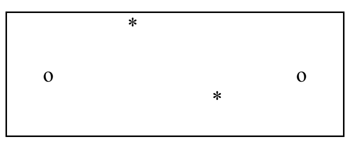


7	8	9	7
4	5	6	4
3	4	5	3
0	1	2	0

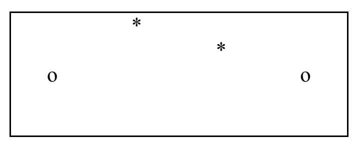
d.



7	9	5	7
4	6	2	4
3	5	1	3
0	2	10	0



7	9	6	7
4	6	3	4
3	5	2	3
0	2	11	0



7	9	8	7
4	6	5	4
3	5	4	3
0	2	1	0

Fig. 21. "Origin-neighbor$_i$-neighbor$_j$-origin" shapes around each triad landmark

---

Lee Humphries is President of www.ThinkingApplied.com, where he devises problem-solving strategies in a variety of fields. Over the years he has had many stimulating conversations about sight-singing with Jack Boyd, Asher Zlotnik, Luigi Zaninelli, Charles Webb, Allen Winold, and his friend Gary Wittlich (with whom he co-authored *Ear-Training: An Approach through Music Literature*). You can contact him at box3@ThinkingApplied.com .