

FOCUS on VOCAL TECHNIQUE

The Singer's Dilemma: Tone Versus Diction?

by Shirlee Emmons

It is axiomatic that singers feel misunderstood by outsiders, that is, by non-singers at large and even by other musicians who do not sing. Not only do singers *feel* misunderstood, they *are* misunderstood. How does this come to be?

The human voice has the capacity--as does no other musical instrument--to express not only music but also drama and literature. For this reason, study of the vocal instrument becomes not just technical and mechanical but also human. To be sure, the vocal cords function on a subconscious level and are best controlled indirectly, often through *imagery*. It is this fact that makes vocal training highly intimate and convinces other musicians that the personal nature of vocal pedagogy is an object of ridicule. ("What does that mean: send it *up and over*?" non-singers say with heavy sarcasm. "Send *what*? Over *where*?")

To be fully rounded artists, singers must use their voices as musical instruments and also as a means of communication. The *musical* use of the voice necessitates a knowledge of how to maintain ideal conditions for a constantly beautiful and musical tone. In contrast, the *communicative* use of the voice requires a command of ever fluctuating symbolic sounds (consonants) that are often noisy. Because singers must do both things at once, the art of singing is essentially contradictory. Herein hangs the tale.

The great voice teacher William Vennard said this: Exactly in the

middle—halfway between the entertainer who disdains music that is “classical,” and the ivory-tower instrumentalist who disdains music that is “programmatic”—sits the poor singer of serious music. Affronted by hearing highly successful entertainers (classified by the public as “singers”) break every rule of good singing and good music, the real singer often does not make peace with the fact that such an entertainer concentrates mostly on diction and dramatic projection, the non-musical elements of the craft. On the other hand, hearing some pure musicologists deride “second-rate” composers like Berlioz for writing descriptive music, they are perplexed. For their music, inextricably wedded to *words*, is by its very nature programmatic. What else could it be?

Vennard bids us digress for a brief look at music history. One sees that voices and instruments were once equal. In the days of madrigals it mattered not whether one sang or played his/her part. Words were engulfed by the contrapuntal themes. Soon, however, voices had to struggle to keep up with the increasing virtuosity of instruments. As voices became successful at this, the words became less important than the display. One syllable often went on for sixteen bars of virtuosic singing, rendering unfathomable and unnecessary the poetic content of the piece. Then Gluck put a stop to all this by means of his reforms. Before long, as composers began to separate what kind of writing was suitable for an instrument and what was best for a voice, long fioratura passages disappeared and the poem became an equal partner with the music. The Lied and the opera—exponents of *dramatic sincerity*—flourished.

Those composers who were the great writers of vocal literature had believed that poetry and drama went hand in hand with music. After a while, this view was no longer held by composers of vocal music, as they struggled to extend the peripheries of compositional techniques. Many began to regard singers and their audiences as a nuisance and an impediment. An electronic sound that imitated a human voice would give less trouble, they believed.

Moreover, the dissonances now primarily used by composers made difficulties for singers, whose singing in tune depended solely upon their ears. Clearly, the singer has no geographical references for pitch: no frets, no valves, no keys, and only a few singers have perfect pitch (the possession of which is debatably *not* an asset for a singer). Despite this, many singers managed to accustom their ears to the new music and excel in its execution. Today, in a development most welcome to singers, new music's latest identity has begun to shift its viewpoint back toward dramatic sincerity, considering music and text to be equally responsible for meaningful vocal music.

Is one element—beautiful tone or intelligible diction—more important?

We see that singing is a paradoxical enterprise. It can flourish only when beautiful sounds issue from the singer's throat, but those beautiful sounds must be accompanied by an illumination of the meaning behind the sounds. Yet the consonants that help accomplish the meaning often pose a real threat to the beauty of the sounds.

The singer's indefatigable quest for a higher level of expression defines the basic elements of singing. They are two: the musical element of the voice (accurate, sustained vowels) and the expressive communication of speech (well-defined consonants). Singers and their teachers seek a diction that is as clear as speech. Truthfully, however, that diction will give only the *illusion* of being the same as speech; it must be quite different in actuality. William Vennard's felicitous phrase explains, "**To sound 'natural' will require *studied artifice*.**" My way of teaching that "studied artifice" is what this article is about.

How to achieve Vennard's "studied artifice"

In my view, the first principle is this: It is very easy to have what is

known as good diction while singing poorly; the real trick is to have good diction while not letting it interfere with good singing. Such skill is not easy to come by. It is not enough to heed the coach's frequent admonition to "just spit it out," or to "sing on the consonants" (Robert Shaw's immortal words), or to "relax; it's just like speaking on pitch." Singing is *not* like speaking. For one thing, experiments have proven that the consonants [*b, d, f, h, s, t, v, m, z*] average .058 seconds each in speech and .108 in song. The semi-vowels [*l, m, n, ŋ*] average .145 in speech and .354 in song. Vowels average .280 seconds in speech and .797 in song. Think of the consequences of these facts. *N.B. You will note that I am using the IPA symbols for the vowels, but not for the consonants. To be frank, one can live without the IPA consonants, but not without the vowels*

For all practical purposes, consonants represent the singer's biggest problem of articulation. The consonants used for purposes of everyday speech are simply not precise enough to be used for singing, especially in American English. Also, remember that the emphasis placed upon a legato line is not a voice teacher's whim; it is crucial to the demands of good singing. This is why the prime tenet of *bel canto* is legato. Yet, the greatest barrier to achieving a legato line is the presence of consonants. Why? Think about it. Consonants close your mouth (which ought to be open most of the time). Consonants tense the tongue (which ought to be relaxed most of the time). Most consonants stop the air flow (which definitely ought to be moving all of the time).

Yet we must allow consonants to sound in order to have intelligible words, one of our singing objectives.... What is the answer?

SHORT but ENERGIZED!

The consonantal movement must be short and rapid, not tentative and extended. In addition, the gesture must be energized, not lackadaisical—in sum, a fast, energized movement of the tongue or lips or any combination

thereof.

For the most part, we singers are not equipped to do such precise movements without training. The sooner singers bite the bullet and learn to do the requisite movements in the proper fashion, the sooner they can hope to become skillful singers. One of my students reported to me that her famous Boston teacher had preached, "You must train that naughty tongue to do your bidding!" This study of consonants cannot be avoided, but it can easily be postponed. Not a good idea for a professional singer.

Now to the specifics. They are onerous and sometimes boring, but absolutely necessary.

Don't move the jaw unless absolutely necessary.

The first skill to be learned is how to snap the tongue rapidly and energetically to the proper area of the palate *without* necessarily moving the jaw, which would of course change the resonance and the beauty of the tone. The consonants most used in the three major languages of singing are

t, d, n, l, and r.

These consonants all touch the palate at the alveolar arch, preferably between the upper front teeth and the arch. (You don't know where that is? Run your tongue from inside bottom of your upper front teeth to the inside top of them. As the tongue leaves the top of the teeth it touches the bony arch running behind the half circle of teeth. This is the alveolar arch.) Try the following to clarify the issue for yourself. Make a medium-sized mouth opening. Leaving the jaw where it is, bring the tongue up to the space in front of the alveolar arch and articulate the consonant *l*. For the five consonants listed above, it is almost not necessary to close the jaw at all (except when *t* and *r* need a little jaw closing because the mouth for the vowel has been very open). To move *only*

the tongue—not the jaw—is a skill that is utterly necessary to learn.

Unless singing in a dialect (as do the characters in Copland's *Tender Land*, for example), refrain from using the stopped *t*. That's the characteristically American one that makes no sound at all. With the stopped *t* the tongue just hits in front of the alveolar arch and releases silently. This, by the way, is the difference between Americans and the British saying, "I saw it." When the Brits say this, one hears the *t*.

Which consonants *do* or do *not* require the jaw to close?

The second skill is to explore exactly where each consonant (other than the aforementioned *t*, *d*, *n*, *l*, and *r*) is located and which muscles do it. Incidentally, remember that the consonants found in the English alphabet are not all that must be mastered. Certainly German and the Slavic languages contain others that you must include in your examination. Remember also that, although the consonant may be spelled differently, the sound is often the same, e.g., *sch* in German and *sh* in English are the same sounds, as in the words *Fleisch* and *flesh*. Dispensing with the arcane phonetic terms (*palato-fricative*, or some such) makes your work a bit easier. Just examine your speech movements and group together in your personal lexicon all consonants that are executed in the same physiological fashion. The following is a practical list:

With *k* and *g* the tongue strikes the middle of the palate. They do not need

a closure of the jaw to execute. (E.g., *cat*, *got*)

B, *m*, and *p* need only a closure of the two lips, *b* and *p* being plosive and *m*

actually sustaining a tone. (E.g., by, my, pie) Although it feels unnatural,

lips together,

but without shutting the jaw, if it is necessary for the tone quality.

W requires a closing of the lips, but without so much pressure. (E.g., witch)

Watch that you do not sing the *wh* (e.g., which) sound in the same way

that you sing the *w* sound. These two words (which and witch) are not

pronounced alike. (It is not *witch witch*.) In the *wh* sound the lips do not

touch. Air is blown between them. These consonants require a closing if

the vowel has been extremely open, but can be executed without closing.

J (or soft *g*) and the *ch* sound require the tongue to strike the palate slightly in

front of the place where *k* does. (E.g., job, gypsy; charm, chop)

The sound for *s* is achieved by raising the front half of the tongue close to

the palate, touching the sides of the tongue to the upper teeth, and blowing

air between the tongue and the palate. *Z* does the same thing, but adds

sound to the blowing air. (E.g., seat, zero) The jaw must close.

The sounds for **sh** (e.g., shall, sharp) and for **zh** (e.g., pleasure, azure) are exactly

the same except for the fact that the **zh** is accompanied by sound whereas

the sound for **sh** is unpitched. The jaw must close.

The sounds for written **x** and **cks** are both in reality a **k** sound followed by an **s**,

actually **ks**. Since it is virtually impossible to pronounce an **s** without

closing the jaw, both consonants must be done by closing. (E.g., six,

sticks) A variation would be in the word asks where the **ks** is preceded by

another **s**, producing **sks**. These combinations are also found in German.

F and **v** are the same consonant, **f** being without sound, and **v** having pitch. They

are both executed by closing the jaw minimally, touching the upper teeth

lightly against the lower lip, and blowing the air through the remaining

space for the **f** but not for the **v**. (E.g., fairy, very)

The **n**, **ny**, and **ng** sounds differ slightly. With **n** (e.g., not) the tongue strikes the

palate just behind the teeth. With the *ny* sound (e.g., *onion*) the tongue strikes in the same place as with *n* and then rolls a little back on the palate for the *y*. For *ng* (e.g., *hung*) a good part of the tongue plasters itself to the palate at about the halfway mark, effectively stopping the tone from going on through the mouth, and sends it on a detour through the nose.

The normal *h* sound (e.g., *happy*) is made in the larynx itself, but the *hu* sound (e.g., *Hugh*, *humor*, *hue*, *huge*) is made by putting the tongue in an [i] position and blowing air through the space, followed by a [u]. Do not confuse this sound with the current popular pronunciation of the four words above (as pronounced: *You*, *yumor*, *yoo*, *yuge*), which changes the *hu* into a simple *y*. This is not only wrong but also vulgar. No closing of the jaw is necessary.

The German *ch* sound after a vowel

(*Note*: If you can't see the IPA symbols below, select this version of the article: [diction.pdf](#) [requires [Adobe Reader](#)].)

In German, the final *ch* of words such as *ich* and *ach* follows the

tongue position

of the preceding vowel. If singers who do not speak German would follow this rule, these particular consonants would not be such a pesky task for them. Moving from front to back of the palate, this is the drill:

ich: put the tongue on the [ɪ] vowel and blow air above the

tongue without moving it.

ech: put the tongue on the [ɛ] vowel and blow air above the

tongue without moving it.

ach: put the tongue on the [ɑ] vowel and blow air above the

tongue without moving it.

och: put the tongue on the Italian [ɔ] vowel and blow air above

the tongue without moving it.

auch: put the tongue on the [u] vowel and blow air above the

tongue without moving it.

Try it. The whole trick is not to move the tongue from the vowel while blowing the air.

Italian double consonants

The flavor of the Italian language comes from the double consonants. Non-Italians must pay close attention to executing these double consonants accurately. To do this is a rhythmic problem. Let's use the word *spaghetti*

spoken in three quarter notes to illustrate. The second quarter note on the syllable *ghe* must also accommodate the double [t] of the last syllable. That is, the double consonant is given a somewhat longer duration than the vowel [ε]. Once the [ε] is articulated, go immediately to the double [t], opening to the syllable [i] exactly on the third quarter note. Effectively, this gives you three equal durational values: SPA GHETT I. The double consonants are not stronger, just longer.

Does it matter to the diction where the consonant is in the word?

Consonants occur in one of three possible positions:

1. initial consonants of the words, especially the first word of the phrase,
2. consonants beginning or ending words in the middle of the phrase,
3. consonants that end the word, especially the last word in the phrase.

1. Initial consonants

When starting a phrase with a word that begins with a consonant, do not prepare the consonant. Instead, prepare the vowel position with your mouth, then inhale, mouth retaining the vowel position. Then, on the beat, snap your tongue, lips, or lips and teeth—with whatever movement the consonant demands—to the consonant, rapidly and energetically, but not in a sustained gesture. (E.g., with the words *Caro mio ben*, prepare the [ɑ] position, inhale in that position, let only the center of the tongue make the *k* sound without changing the [ɑ], and attack the note.) What do you achieve doing it this way? An easier attack, a better tone, less tension, clearer diction. *N.B. See Placido Domingo's Mozart CD for a clear exposition of this technique.*

2. Interior consonants

Those final consonants that occur in the middle of the phrase are always delayed until the very last moment and then attached *to the next word* as swiftly and energetically as possible.

Example, *Joan and Bill stood easily*. (Imagine a quarter note for each syllable.)

Sung: *Joa.....nan.....d[ə]Bi.....ll[ə]stoo....dea...si....ly.*

Note the schwa, [ə]--a very short “uh” sound without much character--inserted after each sounded consonant. Why do we do this? In English, German, and Russian, languages which are famously full of more consonants per square inch than Italian, Spanish, or French, the real problem in deciphering sung words is that (given the obligatory legato) we sometimes cannot tell where one word ends and the next begins. A schwa [ə or ʌ] sounded on pitch (after consonants that have pitch, e.g., *d, b, m, l*) or unsounded (after consonants that have no pitch, e.g., *t, p, k, f*) will separate the two words, such as *an....d [ə]Bill*. Without the schwa, it might sound vaguely like *handbill* or something else. So consonant clusters at the end of one word and consonant clusters at the beginning of the next word must not be bunched together without a schwa. The schwa will help the listener to separate one word from another and to comprehend the sentence. In the case of the unsounded schwa, the empty space between words will have the same effect.

Example, *Both of them took their time*. (Quarter note for each syllable)

Sung:

Bo....tho....v[ə]the....m[ə]too....k[ə]thei.....rti....me[ə].

Note the unsounded schwa between the *k* of *took* and the *th* at the beginning of the word *their*. There is no tone, just barely audible air noise in the schwa,

but the space does the job: it separates the two words so that the listener can comprehend.

Remember: however strange and unnatural this feels and sounds to you, you cannot judge its efficiency until you ask a listener to tell you whether or not he/she heard the word. That is the proof.

The British, it must be said, do not look kindly upon schwas. That is their prerogative. However, I find that their singing without schwas (often somewhat ineffectual) supplies one of the reasons why they tax their voices when singing loudly and sonorously. (Even the Italians, who do not admit the existence of the vowel schwa, adopt schwas when singing in their own language. E.g., on the last page of Alfredo's *Traviata* aria, I have yet to hear an Italian tenor sing the words *in cielo* without executing them in this way: *in*[ʌ]*cielo*.) To do this well, the schwa must be executed (not à la Liza Minelli-- *home* [ɑ..ɑ..ɑ..ɑ..ɑ.] with a lengthy, wide open [ɑ] after the *m* of the word *home*) but as an extremely rapid [ʌ or ə] on pitch. What pitch? The pitch on which the vowel was sung. The pitch will carry; a noise will not.

Listening to pop singers invariably singing a word starting with a vowel as a glottal, we have begun to believe that is the only thing to do. Personally, I am not a fan of this many glottals. My rule is this: if eliding the last consonant of one word to the vowel beginning the second word creates another word, then a prudent glottal will help to clarify. Otherwise, do not do it, unless some unwritten edict makes this maneuver sound un-English or un-American. Use your instincts to make this decision. In general, try not to pepper your singing of English with glottals.

So the rule holds: When the first word ends with a consonant or a consonant cluster and the next word begins with a consonant or a consonant

cluster (e.g., *and strong*), your job is to put a schwa between them (on the pitch of the last vowel of the first word) and move rapidly through the consonant(s) at the beginning of the next word to the vowel of that word. If the second word begins with a vowel (e.g., *and l*), then the final consonant of the first word *and* is attached to the vowel of the second word (*a...ndl*). This accomplishes two important things. We understand the words better, and it is the vowel that is elongated, not the consonant. Thus the singing itself improves.

3. Final consonants

For the very last consonant in a phrase, it is important for you to understand the following principle. Especially in English and German texts, but also true for Italian and French, we do not fully understand any phrase until the last word has been uttered or sung. This means that comprehensibility often hangs on the last word. In order for the audience to understand the phrase, the singer must sing a clear **last word**; in order for the audience to understand the last word, the singer must sing a very clear **last consonant**.

Singers often labor under the misapprehension that it is the first consonant in a word that must be attended to. Coaches, in their effort to help with diction problems, often pursue this route—generally to little effect. For it is not the first consonant in any word that must be totally clear; it is the last one. This is an inviolable principle of the English and German languages.

Thus, the same principles as articulated in section 2., Interior Consonants, apply here. If the final consonant has pitch (e.g., *d, b, m, n, z*, etc.), it should be followed by a tiny schwa on the pitch of the last note. If the final consonant does not have pitch (e.g., *p, t, f, s, k*, etc.), it will be necessary to execute a triple strength spurt of air to accompany the unpitched consonant. Anything less will not carry.

Some other options with consonants, or how to handle consonants on high pitches

For the most part, the lower the pitch for a vowel that comes after the consonant, the easier that note is to execute (a fact that pop singers take total advantage of). The opposite is also true: the higher the pitch for a vowel that follows a consonant, the more difficult it is to sing it, not to mention making it intelligible. That is, on high notes preceded by a consonant one generally has this option: one can be understood, or one can make a beautiful sound. In truth, in a moment defined by a very high note, the word scarcely matters to the listener. The sheer sound of the vocal tone itself carries that moment. Furthermore, it is the composer who must anticipate the problem that will be caused by the word written on the high note. If the composer is well versed in writing for the voice, he/she will do what Mozart did—that is, not write a word that is crucial for the understanding of the plot *only* on that one high note. Instead, he/she will have repeated that word many times before the high note, and reserved the high note moment for tone, not import. Douglas Moore of *Baby Doe* fame once apologized to me for creating such a dilemma for the leading soprano in an earlier opera of his, saying, “If I had known then what I know now, I would not have allowed the audience’s understanding of the drama to hinge on that one word. I would have preceded that high B on the word *kill* (!!!) with several other repetitions of the word *kill* on lower notes. Then you would have been free to just *sing* the B to the best of your ability.”

“Fudging” and/or lengthening the consonant

When faced with the problem of a very high note that is being ruined by a consonant (assuming there have been other repetitions of the word previously or will be repetitions later) the singer can panic, or “fudge” the consonant and modify the vowel, or he/she can do what might be called the

“Pavarotti trick.” At one time in my teaching life, a student of mine was accorded the privilege of working with Pavarotti each morning for several weeks. In the afternoon this student would come to me, and I would ask for the precious details of what The Great Man had said. In this way, I learned that Pavarotti had insisted that the student put the consonant on the lower note that preceded the high note and then go directly to the high note in a great leap of air. He called this “landing *piano* on the high note.” By that, I assume he meant *gently*. And, indeed, it is so easy to accomplish that it does feel like a gentle onset.

It is, of course, a rhythmic problem. For, if you wish to put the upcoming consonant on the preceding note, the time for executing that consonant will have to be subtracted from the preceding note. The vowel of the following high note will have to open right on time. For example, imagine a skip from F to B flat on two quarter notes. Words: *O, God*. Executed in the “Pavarotti” style, it would become *O,GGG.....od*, with a closure of the consonant **G** to the palate on the F and an opening of the vowel [ɔ] on B \flat . In other words, the **G** of *God* would come on the F, and the B flat would have only the vowel [ɔ]. Try it. It is unmistakably what Pavarotti does in his own singing. Remember how everyone admires the clarity of Pavarotti’s “diction.” Listen to a recording. He does “the trick” before every high note. The skip to the high note sounds very easy and gentle. The diction is clear as a bell, because this technique makes the low pitch consonant more audible than if it were put on the higher note.

There is a further plus accrued to the vocal tone when putting the consonant on the lower note before a skip upwards. A plosive which is lengthened will give higher pharyngeal pressure, thus giving the higher note more chest content and, therefore, a darker, stronger quality. The plosives **g** and **k** (center of the palate), **t** and **d** (in front of the alveolar arch), **p** and **b** (lips together), will give, when lengthened and placed on the lower note, a stronger,

darker, more cutting tone. In contrast, the consonants *m* and *n*, when lengthened, will produce a tone lighter and more heady on the following tone. Consider what result you would like to achieve, and choose accordingly.

Lengthening the consonant has another purpose in artistic singing--not just a vocal technical one. That is to produce a more expressive word when desired. For example, in Schubert's "Nacht und Träume" the word *stille* is repeated two times at the end of what might be called the first verse. Lengthening the *sh* sound on the second *stille* creates an expressive, meaningful moment that is very lovely. Extending an *s* or an *f* or any of the various *chs* has the same result. The moment seems to reveal a personal speech pattern and thus appears to heighten the sincerity of the singer. If you want to hear such a technique at work, listen to any song sung by Fischer-Dieskau with the music before you. In adopting this skill, the vowel may begin slightly late. As long as the effect pleases you and doesn't annoy a conductor, go for it.

Back to the issue of "fudging" the consonant before a difficult note. An example would be this. Using *p*, *b*, *m*, *v*, or *f*, all of which must touch the lip to the teeth or the lips to each other, simply do not *touch* whatever is supposed to be touched. Just come close to closing, but don't touch. The listener will intuit what word is intended, especially if the composer has done a good job. (See any of Giuseppe Verdi's arias.)

Context as an aid to diction

Another principle: Context helps us listeners to understand the words. There is no doubt about this. Think of the last time you went to the theater. There is no way that you understood every single word. You guessed successfully at several of them because you comprehended the context.

However, in songs and opera texts, some poetry and prose is written in an arcane fashion and requires time to study to comprehend the meaning—which time is not available to the audience. For example, T.S. Eliot’s poetry is difficult even for those who speak English. Rock poetry or *Woman’s Day* poetry may speak to its audience, but it is not difficult to fathom on first hearing. The language of this poetry is—in that most useful 20th century word—“accessible.” Adding inaccessible text to bad diction will produce a disaster. This means that the audience will probably have a lot of trouble understanding the T.S. Eliot (or other difficult) poetry even when sung with *excellent* diction. To a singer, poetry that is too deep, requiring long and concentrated attention, is a handicap. And also to the composer, and to the poets or authors, many of whom do not wish their words to be used in vocal music for this very reason. They are very aware of this problem. They must deal with it all the time. It is also one of the criteria influencing how the composer chooses poetry to set in his songs. The singer’s problem is to be sure that he/she can illuminate for the audience poetry or prose that is *not* accessible. This equates not only to superb diction but also to performance skills.

Summary

Don’t be put off by the complicated solutions addressed above. In my experience you can better your diction skills within six weeks of real, applied effort. Is it not clear from the necessary length of this article that the singers’ diction problem cannot be mastered simply by “spitting it out?” Far better that singers spend the requisite time to conquer the difficulties and make it their goal to achieve superb diction accompanied by fine singing.

It can be done. Good luck!

©Shirlee Emmons

[Return to Shirlee Emmons' home page](#)

[Return to Vocal Technique page](#)