



by Joel Tall

## HOW TO splice A DIPHTHONG

IF DR. JOHNSON were alive today, and if you were to ask him his estimate of the amount of study needed to learn tape editing, the Great Bibliographer would probably snort derisively and explode, "Why, Sir! It is simply the application of common sense!" I agree with this insidiously fictitious but typical pronunciamento of the immortal Sam, with some reservations. Common sense plus a little observation *will* carry you through, and, in time, you can become an expert tape editor without reading a word of what follows. You may, however, shorten your labors considerably by reading about some of the observations I have made over a period of ten years. I have tried to find reasons for what I have observed, to tie the whole bundle up into a theory of tape editing that, I think, makes sense—common sense, if you like.

An expert tape editor, whether he realizes the fact or not, is interested in many things. He is very much concerned with how he hears sounds, because his hearing is his major editing tool. He senses that a great deal of what we call "hearing" takes place in the brain, and varies

according to the condition of the nervous system. He must know something of speech and phonetics—do people say "a second time" or "a secon' time?" Under what conditions would one version be used and not the other?

The expert tape editor may know nothing of music, but it would help him in his work if he did know the fundamentals, at least of musical nomenclature and notation. He should know how acoustical conditions vary, and the resulting changes in sound. He would profit from widening his knowledge of every phase of sound, hearing, physical acoustics, and psychoacoustics.

Probably the best way to begin to learn tape editing is to learn to recognize speech sounds when played at very low tape speed, that is, at the speed the tape travels when you turn the reels by hand. About all you will be able to recognize at first will be the characteristic "sh" of some "s" sounds. The rest will sound like a series of grumbles, animal noises, and gasps. One must learn to differentiate between these sounds and to know what each sound would be if the tape were traveling at normal speed.



Thus, if you heard a sound that could have been made by a sick lion showing his tongue to the doctor, you should recognize it instantly as the word "I." But that's not all — you should be able to recognize, at low speed, whether that grumble represents a declarative "I," a questioning "I," or some other sort of "I." You should also know, before cutting, whether the "I" is a complete and editable or unfinished and, therefore, uneditable sound. Speech sounds vary according to the succeeding sound. The sound of "I" in the expression "I want" is considerably different from the sound of "I" in "I am." The forming of the mouth in preparation for the following sound considerably affects the sound being pronounced. Speech sounds, in many cases, flow, one into the other, in a blend of the two sounds. The expert tape editor has to know, instantly, whether he can cut such a sound or not. This ability, or judgment, constitutes one real difference between the expert and the tyro.

To acquire this kind of judgment in editing takes a lot of practice — patient and attentive practice. You may find some help, as I did, by studying various works on speech and phonetics. There is quite a bit of valuable information in *Visible Speech*, by Potter, Kopp, and Green (Van Nostrand), but nothing you can read will take the place of unremitting practice in recognition of sounds. You should play a sentence at normal speed and then analyze it, sound by sound, at low speed. Learn to recognize the "hard" sounds — t, b, p, d, and so forth — and the variations in sound of f, v, s, sh, z and similar sound combinations. Then study the various sounds of the vowels and diphthongs. Learn to tell the difference at low speed, of sounds almost but not quite, the same. After a while (how long it takes depends upon you) you will acquire the ability to recognize sounds at low speed.

You might make a game out of listening at low speed. See if it is true that the palindromes "madam," "Hannah," and so forth sound the same backwards as they do forwards. You will find that they do not. About the only sound combination I have found that sounds almost the same played in either direction is the expression "y'see." But this game of playing sounds backwards is more, much more, than a game. It is a valuable way to find exactly where in

the tape one sound blends into another. If you were able to borrow a sound spectrograph, and if you made pictures of the sounds you wanted to edit, you would find one set of meaningful sounds blends into another set of meaningful sounds. There is seldom any definite break, at least not within the same word. But, when you listen to sounds played backwards, you often will find that you can recognize a particular sound more easily. Your hearing is not deceived by the context. You hear only what is there, not what you want to hear, or expect to hear, or imagine that you hear.

I began to use this reverse-listening technique sometime in 1947, quite unaware that there was anything scientific in the method. I was faced with the problem of cutting out the "r" from the word "streamer" to make the word "steamer." Cutting after the "t" sound was child's play, but cutting out the "r" sound completely, without ruining the following "e" sound, was quite a poser. Try it, at 7½-inch-per-second tape speed and I think you will agree with me. The "r" and "e" sounds seemed to blend into each other without a definite change of sound where I could cut the tape. Finally I tried listening to the tape backwards. I could hear the beginning of the "r" sound clearly! I pronounced a few hosannas and went on with my editing.

It seems that others have remarked upon the greater ease of recognition of phonetic sounds when heard in reverse. E. W. Kellogg wrote about it, I discovered later, in the *Journal of the Acoustical Society of America* (Vol. X, pp. 324-326; 1939). W. Meyer-Eppler, of the Institute for Phonetic and Communications Research of the University of Bonn, Germany, wrote about reverse recognition of speech sounds, also in the *Journal of the Acoustical Society of America* (Vol. XXII, No. 6; 1950). He wrote: "It is a well-known fact that the ear may recognize speech sounds by hearing the reversed record, where the context would often mislead it in normal playing, even when we are accustomed to listening phonetically."

You will notice, when you use the above reverse-listening method, that sounds at the beginnings of words, like d, k, and so on, seem less "hard" and not so definitely voiced as in normal "forward-listening." I do not know exactly why this is so. One guess is that the beginning sounds are partially obscured, or masked, in our hearing, when heard in reverse. Another is that, in reverse, some cancellation of certain kinds of distortion takes place, removing some of the transients that may attend the voicing of these sounds. Whatever the cause may be, the method has proved eminently worthwhile.

When you have practiced cutting speech sounds of all kinds and varieties until you wander around mumbling diphthongs to yourself, you may be ready for the next step in editing speech. That is the editing of a talk, sermon, or speech into a shorter version making just as much sense as the original. Here is an area where my own hide is tender from repeated lashings. How many times have I been told by mentors trying to teach me how to write: "When you look over what you have written, and glow with the pride of creation on reading a little gem which you consider the best sentence you've ever composed —

cross it out!" The same thing is true of speeches in general, sales talks, and some lectures. Not only can they be shortened but often they gain both meaning and impact in the process. However, you must observe two rules in editing connected, meaningful speech. You must retain the basic information, both in the literal meaning of the words and in the implied meaning the speaker gives them by his inflections.

The edited version then, if a speech is to retain the character of the original, must retain the mood of the original. It is one of the more artistic aspects of editing that the editor, if he is worth his salt, must sense all the emotions of a speaker correctly in order to interpret his meaning. Sorrow, sarcasm, irony, earnestness, gladness—are all expressed in the way words are spoken. A human voice is capable of expressing fine shades of meaning. These can be ruined by careless editing, even if the *words* of the edited version is correct. Be certain, when condensing a speech, that mood does not change suddenly.

There is nothing that betrays the tyro editor so quickly as the lack of pacing in an edited speech. Every speaker has his own rate of speaking. He will pause periodically, for emphasis, or he will change his pace for some other purpose. He has to breathe occasionally. He will have his own unique speech mannerisms. The edited version must be consistent, in all these respects, with the original. I have always suggested that we edit from the *beginning* of the sound to be cut out to the *beginning* of the sound to be retained—not from the end of one sound to the beginning of the other. If you edit in the way suggested, from beginning of eliminated sound to beginning of retained sound, you cannot help maintaining the speaker's pace between words and sentences. Observe his breathing periods in the original and retain them in the edited version, even if doing so should force you to cut out breathing sounds, from eliminated material, for insertion at the necessary points. Imagine that you are the speaker; before you cut the tape say what you want him to say, to yourself. Try to develop your memory for the way in which phrases were spoken. Above all, edit for naturalness, according to the speaker's character. If the edited speech does not sound as if it could have been spoken that way, it has not been well edited.

Idiosyncrasies of speech identify a person as definitely as his fingerprints. Someone, I cannot remember who, once wrote that of all the leaves in the world no two are exactly alike. I should hate to have to prove that dictum; however, scientists have proved that a voice can be identified by means of sound-spectrograph comparison. Therefore there are probably not two persons in existence who speak exactly alike. There are regional differences in pronunciation, of course, but otherwise the major differences consist of speech habits of many kinds. Some speak in measured cadence; others "sing" their sentences, reaching a peak of emphasis at the midpoint of a declaration. Some speakers sound bare and void of character without their peculiar slurs, "ahems," and "ers." You've got to retain some of these speech sounds to help identify the speaker. Don't attempt, as some editors do, to "clean up" the speech down to the bare bone of the spoken word.

Speech without character is neither interesting nor human; edit so that character is retained.

The spoken word contains meaning far in excess of the written word. In discussing this truism, a trial lawyer in Massachusetts remarked last summer that a high-fidelity binaural recording of what went on at a trial would be, at least to him, of considerable value in retrial of a case in a higher court. He regaled me, or so he thought, with the old story of the man haled into court for stealing a chicken. I'll not bore you with the story. Suffice it to say that the culprit, while not denying, in words, that he stole the chicken, denied it by the way in which he exclaimed, "I stole a chicken!" Accent and inflection made the difference between affirmation and denial. Accent and inflection may convert dull prose into musical speech that is enchanting to listen to. Read a poem by Dylan Thomas and then listen to his recording of the same poem. There is a magical difference. Or read a play that is being acted on some stage, then go and see and hear the play. Accent, intonation, and inflection make the dead words come alive. They also make editing more interesting and more difficult. It is almost impossible to use a heavily inflected word except in the context in which it was spoken. If it should be necessary to edit after a word so inflected, the only way in which it can be done is by a momentary interruption. The interruption may be a cough, a short duration noise of any kind that is in keeping with the scene, or a burst of applause, if that should sound natural. It is possible, in some cases, to have another speaker interrupt at that point.

We have not discussed up to this time a refinement in editing that can be very useful but that requires powers of sound recognition that are acquired only through much practice. I am referring to intercutting, or cutting from one sound within one word to an identical sound in another word. This kind of editing takes a good deal of judgment and finesse. It is a technique that is used occasionally in cutting music. It can be useful in many situations. For instance, suppose you had to change the question "Must I follow?" to the declaration "I follow." The inflection of the questioning "follow" cannot be changed—at least by the methods exposed in this article. The point is, however, that the first part of the word can be used; we have to find a "fo" sound in a

Continued on page 120

## SPLICE A DIPHTHONG

Continued from page 39

declarative tone of voice. Even more important, the "fo" sound must have the kind of accent, or lack of accent, that denotes "end of declaration." When you find the sound you need, cut in the middle of the "f" sound of both "follow" and "fo" and splice them together. If you've got good recognition and judgment, the word you have constructed will be perfectly credible and natural sounding.

This technique can also be employed in correcting embarrassing mispronunciations and stumbles on the part of a speaker, which he corrects suddenly, with consequently heavy accentuation. Be careful, however, not to create an error of pronunciation yourself in cutting out stumbles. Suppose a speaker said "momento—er—memento" and you wanted to edit the error. The accent of "memento" indicates it is a correction, and it is incorrectly accented, no matter how much you may edit it. Shall we let this speaker stew in his own error? Or shall we cut out a little of the over-accented "e" sound in "memento," substitute it for the "o" sound in "momento" and listen to the result for naturalness? The way you exercise your judgment in a case like this is a good barometer of your ability to edit speech. In this case it might be best to remove all but the beginning part of the "o" sound, which, since this "o" is unstressed and generally elided, would probably make the word sound perfectly natural as "memento."

