Contents

CHAPTER 2

The Human Voice

INTRODUCTION

One of the first things that a newborn baby does is to cry in a wonderfully loud, natural, unimpeded, and open manner. This most basic activity provides a means of communicating with those around to request essential needs for living and growing up. A baby is born with the ability to make sounds using the voice production instrument, and during the early years of childhood, the sounds produced develop into the language in local use through listening, imitating, and observation of the responses obtained.

The cry of a newborn baby is acoustically efficient, free, and very well projected. It can be heard at a considerable distance and the instrument itself is working in a highly efficient manner. However, this natural formula for playing the vocal instrument is rarely left unhindered and therefore does not last. A child's developing use of the vocal instrument is conditioned by parental and peer response, which so often serves to inhibit efficient voice production and dampen vocal performance confidence as the child is told for example, to "pipe down," "stop making such a noise," "stop shouting as we cannot hear ourselves think," "only speak when spoken to," "be seen and not heard," "make a noise quietly," or "stop singing" Depending on the severity with which such instructions are given, a typical response might be some degree of clamming up in terms of vocal output. This is likely to be accompanied by and in part due to increased muscular stress, especially in the neck and shoulder regions, which is in itself a common habit associated with 21st century living that is fundamental to poor vocal health. Coupled with the psychological result of being told by parents and peers that a loud vocal output executed in a free and efficient manner is not routinely acceptable in daily life, the clamming up and stressed response becomes normal vocal habit for many.

Unpacking these and other aspects of less efficient, unnatural, and clammed up vocal output is a key step to healthy voice production and being able to project the voice efficiently in a babylike fashion. Some knowledge of the voice

production process in terms of its anatomy and physiology, as well as the acoustics of both the voice production process and rooms, can greatly enhance progress in healthy voice development. Singing and speech make use of the same vocal instrument, and the underlying anatomical, physiological, and acoustic principles involved are common to both activities.

This chapter provides an introduction to the anatomy and physiology of the human vocal instrument and then focuses on the resulting acoustic output. The presentation concentrates only on those aspects that are vital for a proper and fully informed understanding of the basics of human voice production in the context of vocal health and efficiency, as well as vocal production in different acoustic environments and making a successful voice recording. The singing as well as the speaking voice is considered, and the chapter ends with a number of everyday tactics that can be employed to maintain the voice in a healthy state.

VOICE PRODUCTION

This section describes the main parts of the human body that are involved in voice production, whether speaking or singing. In order to produce a sound with any system, whether an acoustic musical instrument, an environmental noise, the call of an animal, or an electronically synthesized sound, three essential features must exist:

- power source
- sound source
- sound modifiers

Human voice production during speech or singing is no exception, and therefore it will be described here in terms of the *power source*, *sound source*, and *sound modifiers*. During sung notes, these correspond anatomically to the action of the *lungs*, the *vocal folds*, and the *vocal tract*, respectively. These are illustrated in Figure 2–1, which also shows an equivalent mechanical model, indicating with double-ended arrows those parts that can be moved during speech or singing. The relevant anatomical and physiological detail of the power source, sound source, and sound modifiers are described below in terms of their function during voice production, the acoustic output, and useful tips for maintaining a healthy voice.

Power Source in Voice Production

The power source in a musical instrument might, for example, be the bow moving across the string of a stringed instrument; a finger plucking a string on a stringed instrument; the lungs blowing air into a woodwind or brass instrument; a finger striking the key of a piano or harpsichord; the electrical power supply of an electronic instrument; the blower of a pipe organ; or the stick or beater striking a percussion instrument. For the human voice, the power source is the flow of air from the lungs via the throat and mouth and/or nose during exhalation. Indeed, it is the same power source used when playing woodwind and brass musical instruments.

Breathing is a natural function which is automatic and basic to life itself. While the airways are open and the lungs maintain a higher air pressure than the
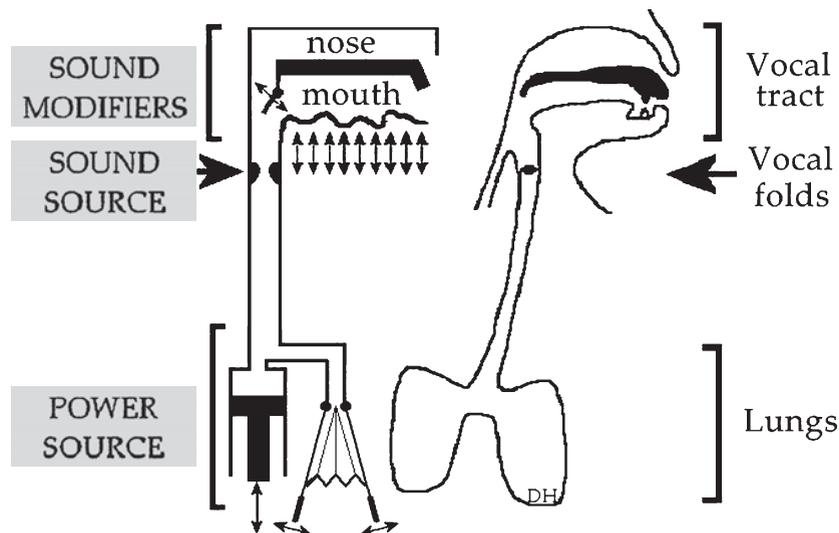
Figure 2–1. *An overview of the human vocal instrument when the vocal folds are vibrating during either singing or speech in terms of its three main constituent parts: power source (lungs), sound source (vocal fold vibration), and sound modifiers (vocal tract spaces). The anatomical equivalent is shown on the right and an equivalent simple mechanical model on the left.*

atmospheric pressure outside the body, air flow is sustained from the lungs to the outside world via the mouth and/or nose. When lung air pressure is lower than the atmospheric pressure of the local environment, air flow is sustained to the lungs from the outside world. This is the basic physics behind how we breathe. We change the air pressure in the lungs relative to atmospheric pressure when we breathe. On breathing in when singing or speaking in a healthy manner, air is drawn into the lungs by enlarging the lung spaces through muscular action, which is equivalent to pulling the piston in Figure 2–1 downwards and the bellows outwards. This produces a lung pressure that is lower than external atmospheric pressure (same quantity of air now occupying a larger volume results in a reduction in

pressure), and air will flow into the lungs (providing the upper airway is open). When we breathe out, through muscular action we contract the lungs (equivalent to pushing the piston in Figure 2–1 upwards and/or the bellows inwards), thus producing a lung pressure that is higher than external atmospheric pressure (same quantity of air now occupying a smaller volume results in an increase in pressure) and air flows out from the lungs (providing the upper airway is open).

The lungs themselves are such that if they were removed from the body they would shrink greatly in size. Each lung can in this respect be considered, albeit rather crudely, as being somewhat similar to a balloon. However, there is a fundamental difference between a lung and a balloon in terms of inflation and

deflation. The lungs are supported externally within the rib cage and from below by the diaphragm so that they can be physically enlarged to suck air in. Breathing in and out is a result of lung expansion and contraction, which is achieved by the actions of muscles as illustrated in Figure 2–2. First, there is a group of muscles that can move the rib cage by expanding it outwards or contracting it inwards. The muscles that join with and control the size of the rib cage during breathing are known as the intercostals. The inspiratory intercostals expand the rib cage and are therefore used when breathing in, and the expiratory intercostals contract the size of the rib cage and hence can be used when breathing out.

Second, there is the action of the diaphragm which is attached to the lungs. The diaphragm is bowed upwards below the lungs when it is relaxed, as shown in Figure 2–2. When it is contracted, it becomes shorter and its shape flatter, expanding the lungs by pulling them downwards (like a piston in a cylinder). In addition, the lower rib cage is opened outwards (rather like a blacksmith's bellows). The lower part of Figure 2–2 illustrates an ideal breathing sequence for which there is a flip book version in Appendix 2. The diaphragm sits over the abdominal wall, and since
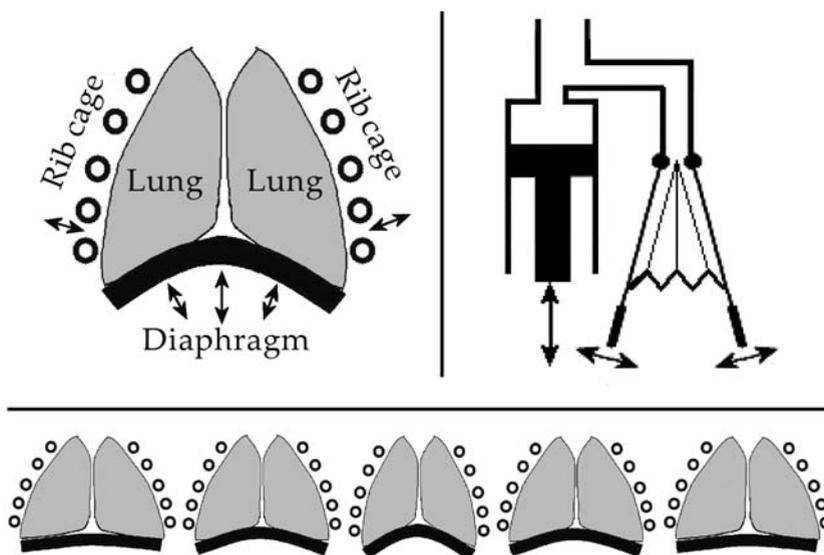
Figure 2–2. *An overview of breathing during singing and speech showing how the lungs can be expanded and contracted using the rib cage and/or the diaphragm, alongside the power source part of the equivalent mechanical model shown in Figure 2–1. The double-ended arrows indicate movement which can increase and decrease the volume of the lungs, and the model indicates the piston- and bellows-like nature of lung action during breathing. The lower part of the figure shows an idealized breathing sequence for which a flip book version can be found in Appendix 2.*

the volume of the abdomen itself and its contents cannot be altered appreciably, any diaphragm contraction serves to push down on the abdomen, which causes the abdominal wall to bulge outwards and air to enter the lungs (if the airway is open). The diaphragm is relaxed following contraction, and it returns to its rest position and air is expelled from the lungs (if the airway is open). Note that the lungs are not empty when the diaphragm is at its rest position; they can be further compressed to enable longer phrases to be spoken or sung. Abdominal wall expansion and contraction are readily observed externally in the region of the navel, and this provides a useful indicator of diaphragmatic breathing.

In summary, during breathing the following muscles can be used:

■ breathing in: the inspiratory intercostals and/or the diaphragm
■ breathing out: the expiratory intercostals and/or the abdominals

Notice that these are the muscles we *can* use to breathe and stay alive. The upper chest region can also become engaged in the process of breathing, as observed for example during rapid panting. However, healthy voice use requires that the upper body (chest, shoulder, and neck region) remains relaxed in order that the neck and larynx are relaxed and under no excessive strain. This precludes the use of the upper chest region for breathing when engaged in healthy voice production, so the predominant muscles used for breathing are the diaphragm, intercostals, and abdominals. Many voice teachers refer to the notion of *support* or *supported breathing*, which provides a practical form of instruction, something

that might be termed a *psychological hook*, to focus the mind of the performer on controlling (or supporting) the lungs from below.

Sound Source in Voice Production

The sound source during sung notes results from the vibration of the vocal folds in the larynx. In this book, the term *vocal folds* is used to describe the vibrating elements in the larynx, because it describes most appropriately the physical nature of the vibrating structures. In medical circles, the vocal folds are more usually referred to as the *vocal cords,* which has its origins in what Manuel Garcia saw when he looked down the throat with his 45 degree mirror (probably seeing the light reflecting from the upper edges of each vocal fold, which would have appeared to be string- or cordlike). The media and other sources often use the term *vocal chords,* which is a misnomer, although one quick-thinking York student justified the use of this term in his submitted work by arguing that the observed shape of each vocal fold from above was a *chord* of a circle!

When a sung note is produced and the vocal folds vibrate, the resulting sound is heard as having a *pitch*. Such sounds are described as being *voiced,* because they involve vocal fold vibration in the larynx or *voice box*. Not all of the sounds used in spoken communication are voiced and produced as a result of the vibrating vocal folds, however. There are *unvoiced* or *nonvoiced* sounds in speech that result from air being forced past a narrow constriction in the mouth or oral cavity, such as the final consonants in the words *pass*, *stiff*, and *pitch*.

Finally, there is a third sound source used in voice production which is a mixture of the voiced and voiceless source, when the vocal folds vibrate and air is forced past a constriction, and this sound source is termed *mixed*. The final consonants in the words *fez*, *pave*, and *badge* have a sound source that is mixed.

In summary then, there are three sound sources used in speech and singing as follows:

■ **voiceless** (involving air being forced past a constriction in the vocal tract)
■ **voiced** (involving vocal fold vibration)
■ **mixed** (involving voiced and voiceless sound sources)

In terms of vocal training, almost all effort relating to the sound source is devoted to voiced sounds and the vibrating vocal folds. This is particularly true for singing training, since a basic requirement is to gain a much wider pitch range than is used for normal speech.

A voiceless sound source involves a narrow constriction somewhere in the vocal tract, for example between the upper teeth and lower lip during the production of the final consonant in *stiff*. If air flows sufficiently fast through the constriction, it becomes *turbulent* and a noiselike sound is produced. Such sounds are known phonetically as *fricatives*. The occurrence of a noiselike sound when air flow is rapid enough can be confirmed by forming a constriction between the upper teeth and lower lip in preparation for producing an *f* sound and adjusting the air flow from slow to fast while listening to the acoustic result. Voiceless sounds have no definite pitch associated with them; you cannot sing notes on them—try singing the final consonants in *pass*, *stiff*, and *pitch*. The production of

a voiceless sound source is learned as speech is acquired, and unless there is some speech-related issue about inappropriate positioning of the constriction within the vocal tract, which would be dealt with by a speech and language therapist, nothing additional is needed for the professional voice user. The remainder of this section therefore concentrates on the voiced sound source.

The voiced sound source results from the vibration of the vocal folds in the larynx. The larynx is situated in the neck, and it can be located by moving the side of an index finger gently up and down the front of the neck to find the prominence on the thyroid cartilage, known as the *Adam's apple*, which can be observed in the illustration of the larynx shown in Figure 2–3. The Adam's apple is usually more obvious and clearly visible for men than it is for women, because the adult male larynx is approximately twice as large in its linear dimensions. If the side of the index finger is placed in contact with the neck on the prominence of the Adam's apple while swallowing, a vertical movement of the whole larynx structure can be felt. This demonstrates that the larynx is supported by muscles in the neck; it is not held rigidly in position.

Vocal fold vibration for a voiced sound source is initiated by bringing the vocal folds closer together horizontally—a movement known as vocal fold *adduction*. Voiced sounds are normally produced when exhaling (breathing out). As air is expelled past the gap between the adducted folds (this gap is known as the *glottis*), the velocity of air flow must increase because the airway is narrower. One physical consequence of increasing the velocity of air flow due to a constriction is that the push or pres-

sure that it exerts on the sides of the tube is reduced. This is known as the *Bernoulli effect*. It should be noted that it is possible to produce voiced sounds when inhaling (breathing in). This is something that can occur automatically when one is communicating in a high state of panic, shock, or fright to allow communication to take place continuously, even when breathing in. Producing a voiced sound source when inhaling is also part of some vocal warm-up/cool-down exercises.

The Bernoulli effect is also the principle upon which aircraft fly. Aircraft wings are shaped as shown in the upper part of Figure 2–4. Air flowing across the upper surface has further to travel due to the upward curve in the wing profile, and therefore less pressure is exerted downward on the upper surface of the wing compared to the pressure exerted upward on the lower surface, resulting in *lift*, as illustrated. The Bernoulli effect as it relates to the closure of the vocal folds can be demonstrated by blowing across a sheet of paper held at the end nearest the lips, as shown in the lower part of Figure 2–4. The sheet will rise up (note the similarity in shape between the curved sheet and the upper surface of the aircraft wing) due to the Bernoulli effect.
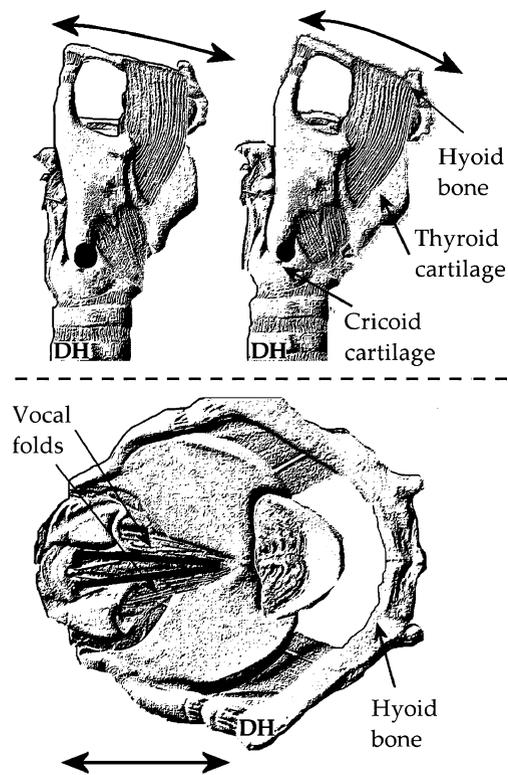
Figure 2–3. *Illustration of the tilting mechanism of the larynx, which enables the length of the vocal folds to be altered and thereby the fundamental frequency of their vibration and the perceived pitch. The upper panel shows a side view to illustrate how the thyroid cartilage hinges on the cricoid cartilage (marked with the black circle). The lower panel looks down on the larynx, revealing the vocal folds, which are stretched and relaxed as a direct result of the tilting mechanism.*
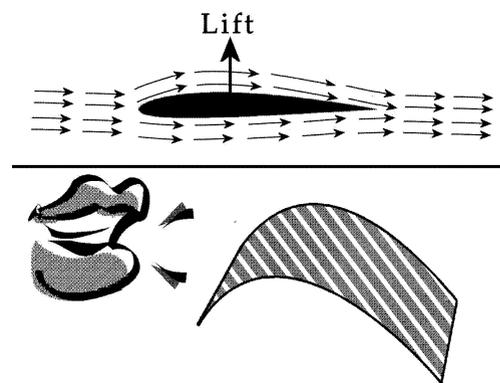
Figure 2–4. *Illustration of the Bernoulli principle as it relates to how aircraft fly based on the profile of their wings (upper), and a demonstration of the Bernoulli principle by blowing across a sheet of paper held as shown at the end nearest the lips (lower).*

During vocal fold vibration, air flows between the adducted vocal folds through the narrowed glottis, where the air velocity increases with the consequential decrease in pressure on the sides of the tube (the vocal folds themselves), as described by the Bernoulli effect. The reduction in pressure on the vocal folds acts to move them towards each other, in a manner somewhat analogous to the lift on an airplane wing. The result of moving closer together is that the glottis is narrowed even more, the air flow velocity increases, the pressure exerted on the vocal folds (tube walls) reduces, and the force pulling the folds together increases. The vocal folds therefore accelerate towards each other as they get closer together, until finally they meet at the midline with a "snap" as the glottis closes.

From the closed position, the vocal folds will open because they have closed off the air flow from the lungs, where air is under pressure. In addition, the folds have a natural tendency to return to their rest/starting position—each vocal fold can be thought of as behaving like an oscillating pendulum. Each vocal fold will move like a pendulum past its rest, or *equilibrium*, position, on to its fully open configuration, and back towards its equilibrium position. The Bernoulli effect again comes into play and the cycle repeats, resulting in sustained oscillation. As the vocal folds vibrate, their lower edges will meet and part before their upper edges, since the folds have depth as well as width (see Figure 2–5), and the Bernoulli effect acts on their lower edges first, due to the direction of air flow. A flip book version can be found in Appendix 3.

In speech and singing, the pitch of the voice is always changing. Even in

Closed phase   Open phase
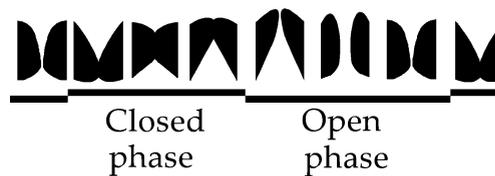
Figure 2–5. *Illustration of the sequence of vibration of the vocal folds viewed from the front (a flip book version of this figure is available in Appendix 3). Notice that opening and closing start from the lower margins and move upwards.*

singing when one attempts to sing a steady note, there will be small changes in pitch. During speech, changes in voice pitch are the "tune" of the language, or *intonation* pattern. English uses intonation to signify, for example, whether or not one is uttering a statement or a question as in the following: *"That train is late!"* and *"That train is late?"* Singers change the pitch to alter the note they are singing, and to tune their voices with other singers or any accompanying musical instrument(s). In speech, pitch tends to be thought of in terms of a changing contour, whereas in singing, pitch relates to discrete notes.

The pitch of the vibrating vocal folds can be changed by altering their mass, tension, and/or elasticity; this is described by the *myoelastic aerodynamic theory of vocal fold vibration* (see Van den Berg [1958] in the further reading list for more details). Increasing the mass, reducing the tension, or making the elasticity smaller will have the effect of lowering the pitch, and vice versa. In practice, the mass can be changed by holding a portion of each vocal fold immobile, which means that their vibrating masses are reduced and the pitch will rise. The vocal folds are supported within the larynx within a hinged structure, as illustrated

in Figure 2–3, in such a way that the folds can be stretched and released, raising and lowering their f0, respectively.

Sound Modifiers in Voice Production

The acoustic characteristics of a sound will be modified by the spaces through which it passes in much the same way as the sound of the voice varies in different rooms and buildings. In the case of speech and singing, the sound modifiers are the spaces through which the sound source passes to emerge from the between the lips and/or nostrils of the speaker or singer. It is the shape of these spaces that serves to modify acoustically the output from the sound source. There are two spaces that make up the vocal tract:

- **the oral cavity** (the space between the glottis and the lips)

- **the nasal cavity** (the space between the velum and the nostrils, or the nose)

The main way in which the shape of the oral cavity (mouth and pharynx in Figure 2–6) can be altered is by moving the tongue, jaw, and lips. Such moving parts are known as the *articulators*, and those that can be moved when speaking or singing are illustrated in Figure 2–6 with double-ended arrows on the equivalent mechanical model of the sound modifiers.

The main articulators serve to alter the shape of the mouth, and the main ways in which this can be achieved in speech and singing are by varying the height of the jaw, the position of the lips between being rounded (as in the vowel in *boo*) and spread (as in the vowel in *bee*), and/or by changing the shape of the tongue by increasing the constriction between it and the hard palate using its
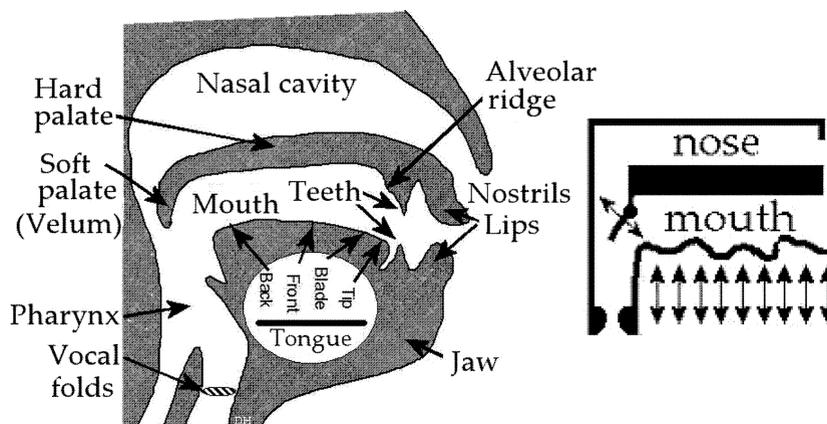

Figure 2–6. *The human vocal tract showing the main parts associated with voice production (left) alongside the sound source and sound modifier equivalent mechanical model (right), shown in Figure 2–1. The parts that can be moved are indicated by double-ended arrows on the mechanical model and their names can be found by referring to the vocal tract (left).*

tip, blade, front, or back. The nose is rather different because its shape cannot be altered. It is used in voice production for sounds such as the final consonants in *boom*, *bean*, and *bring*, the so called *nasal* consonants in English. The nose is engaged by changing the position of the *soft palate* or *velum*, which works as a valve to allow sound to pass through the nose or not, depending on whether it is lowered or raised, respectively (see the mechanical model in Figure 2–6). It is possible to feel the action of the velum if a hum is produced that is broken up by forming but not releasing a *b* (the first consonant in *boo*); one might write this as *mmmbmmmbmmmb*. "Not releasing a *b*" sound means not opening the lips—let it revert to the hum. It should be possible to feel the action of the velum as it is raised when the *b* is formed (to shut the nose off from the airstream), and lowered for the hum (to allow air to flow through the nose and out via the nostrils).

The minimum set of sounds required to distinguish the words of a language are those that are uniquely in the words of that language. For example, the English words *ton*, *done*, *shun*, *son*, *gun*, *run*, *won*, *nun*, *fun* indicate that the initial consonant sounds are unique phonemes for English, since exchanging them in this context produces different meaningful words. Similarly, the vowel sounds that distinguish the words *bat*, *bit*, *but*, *bet*, *boat*, *bait*, *bite*, *bought*, *beat*, *boot*, *Bert*, and *Bart* are also phonemes of English. It turns out that English has 24 consonants and 20 vowels (a total of 44 phonemes) when its unique sounds are considered, which is very different from the 5 vowels and 21 consonants (a total of 26 letters) that exist in the alphabet used when writing words; the correspondence between the phonemes used to indicate how a word is spoken and how the word is spelt is rarely one-to-one.

The articulation of the phonemes in English would be described by phoneticians in terms of three descriptors: *voice*, *place*, and *manner*, which indicate whether the vocal folds vibrate or not (voice), where (place), and how (manner) the sound is produced. Table 2–1 lists the 44 phonemes of English using the SAMPA (Speech Assessment Methodologies Phonetic Alphabet) transcription system introduced by John Wells in 1989. The SAMPA system is used here because it makes use only of characters that are available on a standard computer keyboard using ordinary fonts. For each phoneme, an example word (from the world of yachting) is provided along with its SAMPA transcription, and for the consonants, their voice, place, and manner labels are provided, which are described in the next sections.

Voice

The voice label indicates whether or not the vocal folds vibrate during the production of the phoneme, which is described as being either *voiced* (V+) because the vocal folds vibrate or *voiceless* (V–) because they do not. A quick check to confirm whether the vocal folds vibrate while producing a sound can be made by either (a) putting hands over the ears and listening for a loud buzzing sound, (b) trying to sing the sound, or (c) feeling either side of the throat gently near the level of the Adam's apple for vibration. A number of English phonemes differ only by voice including the initial consonants in *Sue* and *zoo*, *fire* and *via*, *chew* and *Jew*, *pan* and *ban*, *ton* and *done*, as well as

Table 2–1. The 24 consonants and 20 vowel sounds in English with their SAMPA symbols (Sym.), example words (Word), and SAMPA (see Wells, 1989, in the reading list) transcription (Trans.). The voice, place, and manner descriptions are listed for the consonants

ENGLISH VOWELS

Sym.	Word	Trans.
i	neap	/nip/
I	jib	/dZIb/
E	red	/rEd/
{	anchor	/{Nk@/
A	hard	/hAd/
Q	locker	/lQk@/
O	port	/pOt/
U	foot	/fUt/
u	food	/fud/
V	rudder	/rVd@/
3	stern	/st3n/
@	tiller	/tIl@/
eI	weigh	/weI/
aI	light	/laIt/
OI	oilskin	/OIlskIn/
@U	row	/r@U/
aU	bow	/baU/
I@	pier	/pI@/
E@	fare	/fE@/
U@	fuel	/fU@l/

ENGLISH CONSONANTS

Sym.	Word	Trans.	Voice	Place	Manner
p	rope	/r@Up/	V–	bilabial	plosive
b	buoy	/bOI/	V+	bilabial	plosive
t	tide	/taId/	V–	alveolar	plosive
d	deck	/dEk/	V+	alveolar	plosive
k	cabin	/k{bIn/	V–	velar	plosive
g	galley	/g{li/	V+	velar	plosive
T	thwart	/TwOt/	V–	dental	fricative
D	weather	/wED@/	V+	dental	fricative
f	fog	/fQg/	V–	labio-dental	fricative
v	rung	/rVN/	V+	labio-dental	fricative
s	sea	/si/	V–	alveolar	fricative
z	zenith	/zEnIT/	V+	alveolar	fricative
S	ship	/SIp/	V–	palato-alveolar	fricative
Z	treasure	/trEZ@/	V+	palato-alveolar	fricative
h	heeling	/hilIN/	V–	glottal	fricative
tS	chain	/tSeIn/	V–	palato-alveolar	affricate
dZ	jibe	/dZaIb/	V+	palato-alveolar	affricate
m	mast	/mAst/	V+	bilabial	nasal
n	main	/meIn/	V+	alveolar	nasal
N	rigging	/rIgIN/	V+	velar	nasal
w	winch	/wIntS/	V+	bilabial	semivowel
r	rain	/reIn/	V+	alveolar	semivowel
l	lee	/li/	V+	alveolar	semivowel
j	yacht	/jQt/	V+	palatal	semivowel

cot and *got*. In each of these examples, the first of the pair is voiceless and the second is voiced. Table 2–1 shows the voice label for each English consonant.

Place

The place of articulation describes where in the vocal tract there is either a complete closure or vocal tract narrowing. The main places of articulation used for English consonants are shown in Figure 2–7 and are listed in Table 2–1. The *bilabial* sounds /p/, /b/, and /m/ involve contact between the lips. The *labio-dental* sounds /f/ and /v/ result from contact between the lower lip and the upper teeth. *Dental* articulation is used for the sounds /T/ and /D/, which use contact between the tongue tip and the upper teeth. The *alveolar* sounds are /t/, /d/, /n/, /l/, /s/, and /z/, for which the tongue tip or blade makes

Figure 2–7. *The places of articulation used for English consonants.*

contact with the alveolar ridge. The /r/ sound is *post-alveolar* because it is usually produced with tongue contact further back along the hard palate than for the alveolar sounds. There is close approximation between the front of the tongue and the area between the alveolar ridge and hard palate for /S/, /Z/, /tS/, /dZ/, and /j/, which are therefore known as *palato-alveolar*. For /k/, /g/, and /N/, the back of the tongue makes contact with the soft palate or velum and their place is described as *velar*. The sound /h/ is produced with a close approximation of the vocal folds, and its place is known as *glottal* (the space between the vocal folds is the glottis).

The vowel sounds in British English include both those that remain steady throughout (/i/, /I/, /E/, /{/, /A/, /Q/, /O/, /U/, /u/, /V/, /3/, and /@/), which are known as *monophthongs*, and those that change from one vowel to another during their production (/eI/, /aI/, /OI/, /@U/, /aU/, /I@/, /E@/, and /U@/), which are known as *diphthongs*. The production of monophthongs is described in terms of four elements: (a) how *close* or *open* the constriction is between the closest part of the tongue to the roof of the mouth, (b) whether it is the *front*, *center,* or *back* of the tongue which is making that constriction, (c) whether the lips are *rounded* or *unrounded*, and (d) whether the vowel is *nasalized*.

Traditionally, vowels are shown on a *vowel quadrilateral*, which shows the position of each vowel by indicating the position of the highest point of the tongue. The vowel quadrilateral for the English monophthongs is shown in Figure 2–8, along with an indication of the position of the quadrilateral itself in relation to tongue position in the mouth.

Figure 2–8. *The approximate position of the vowel quadrilateral within the vocal tract (upper left) with vowel quadrilateral plots of English monophthongs (upper right), closing diphthongs (lower left), and centering diphthongs (lower right) plotted on vowel quadrilaterals. SAMPA (Wells, 1989) symbols are used throughout.*

The vertical and horizontal axes are open/close and front/back, respectively. Lip rounding can be externally observed. The full description for the vowel /i/ is front, close, and unrounded, while /u/ is back, close, and rounded, and /{/ is front, open, and unrounded.

Any vowel can become nasalized when the velum is lowered and the nasal cavity is coupled to the oral cavity. This happens when there is a nasal consonant either before or after the vowel, and the velum is either still lowered after the nasal consonant, or lowered in preparation for the nasal consonant. A nasalized vowel is sometimes described as

sounding *hollow* when compared to its non-nasalized counterpart.

Diphthongs are vowels which do not remain steady; rather, there is a change from one monophthong to another. This can be readily appreciated in Figure 2–8 where the traditional representation of each diphthong is shown as an arrow from the first to second monophthong, which provides a representation of the associated movement of the highest part of the tongue during their production. For example, the diphthong /aI/ (*eye*) involves a glide between the monophthongs /a/ and /I/, and the diphthong /I@/ (*ear*) glides between /I/ and /@/.

been specifically acoustically designed as performance spaces. In many cases, humans are using their voices on a daily basis in rooms that have been constructed with no consideration having been given to acoustic design, such as many school classrooms, seminar rooms, lecture halls, hotel ballrooms, club back rooms, and church halls. Intelligibility depends on the relative levels of the direct sound and the reverberant field; this and the importance of the critical distance have been introduced in the section Sound Modification by a Space. In terms of a vocal performer, there are two key aspects to consider: (a) how well the performer can monitor her/his own voice, and (b) how comfortably and intelligibly the audience can hear the sound. There are a number of things that can typically be used to improve both these situations in practice, and practical suggestions supported with underlying reasoning are provided below.

PERFORMING TO BEST ACOUSTIC ADVANTAGE IN A SPACE

Introduction

Any performance requires preparation, and since the acoustics of the performance space affect the sound received by the listeners, attention should be paid to achieving the best acoustic result from the space itself. Some time needs to be devoted to this prior to any performance in order to gain familiarity with the acoustics of the space itself, and to consider making changes to its layout. No special equipment is needed but the results can be dramatic in terms of the

overall performance as seen and heard by the audience. The acoustics of the space should be viewed as a tool that is available to the vocal performer. While proper attention to making the space work well acoustically can enhance the final result, using it as offered having gained no familiarity with it in advance can trip up even the best vocal practitioners to the detriment of the final result.

Bearing in mind that the acoustics of a space are determined by the relative positions of the sound source and the listener, it is worth thinking creatively about how a room is laid out for a performance. Even if there is a stage, it is not necessarily the case that that stage is the acoustically optimal place from which to perform. If the audience seating is not fixed, the possibility of moving the seating around should be considered. This is especially important when working in spaces that have not been designed acoustically for optimum vocal performance such as most church halls, school classrooms, seminar rooms, conference breakout rooms, hotel ballrooms, and other spaces where the seating is moveable. When considering where to perform from within a space, however, there are clearly also visual and practical issues to resolve when considering changing the sound source and/or listener position for best acoustic effect.

Different types of music work better acoustically in different spaces, and choral conductors should be aware of this when planning programs to be performed in different venues. Music with much rapid contrapuntal detail will tend to become blurred and the detail lost in a building with a high reverberation time, whereas slow-moving polyphonic music works can be significantly enhanced when performed in spaces with high reverberation times, such as a cathedral.

Acoustic, Visual, and Practical Considerations

Acoustically it is important (see section on Sound Modification by a Space) that (a) the balance between the direct sound and the reverberant field is optimized for intelligibility, pleasantness, and overall loudness, and (b) the performer is able to monitor the output comfortably to combat any tendency towards vocal strain. It is better acoustically to have the audience close to the performers to allow a maximum number to be within the critical distance, thereby providing them with a high direct sound level. If those in the audience who are further away are having difficulty hearing a vocalist, then the acoustic level can be raised by increasing the reverberant field level. This requires an increase in reflected sound, which might be achieved by opening curtains to acoustically expose window glass, or by removing acoustically absorbing surfaces or objects. Conversely, if the sound is too reverberant, which will tend to impair intelligibility particularly for those at the greatest distance from the performers, the reverberant field level can be reduced by closing curtains and/or adding acoustically absorbing surfaces or objects. Such modifications have to be done judiciously, however, to achieve a balance between overall level and intelligibility, noting that this can never be optimal for every member of the audience.

Visually it is important that the performers can be seen clearly by the audience and that the number of blind spots (such as seats with obscured views due to pillars, other members of the audience sitting further forward, or other obstructions) is minimized. It is also essential that the performers can see the audience easily to gain and maintain eye contact, to enable their reaction to be gauged during the performance, and to make appropriate changes as the performance progresses. There may, of course, be issues relating to the lighting to be resolved to ensure that visual contact is two-way. If there are blind spots, it can be useful if the performer moves around whenever possible to help resolve them, but it is worth being aware that movement is likely to create new blind spots elsewhere. A constant awareness of one's audience gained by looking around will keep track of and allow something to be done about full audience inclusion. Bear in mind that a blind spot blocking the view for a member of the audience is also a blind spot preventing the performer from seeing that member of the audience.

Practically, there can be a large acoustic difference in performing from different positions. In many situations it is not feasible to move the audience; perhaps the seating is fixed, there is a stage, podium, or lectern, or regulations state that the fire exit doors have to be at the back, front, or side. However, as a performer there is always the possibility of making an informed decision as to the performance position. The presence of a lectern or podium does not necessarily indicate the optimal position from where to perform; its position has more likely been decided based on overall visual appearance.

Working the Space to Best Acoustic Advantage

Finding a good acoustic position from which to perform can be done by testing the acoustics of the space by moving around it using a handclap as a sound

source and listening to the result. A hand-clap is a single, short acoustic sound input to the space, and the sound modifying effect of any space will present the direct sound, early sound, and reverberant sound (see Figure 3–4) to any ear or microphone within the space. One thing to listen out for is flutter echoes, which are best described as a *ping*—a ringing note. If flutter echoes are present, it means that any sound produced by a sound source placed in the position where the handclap was made will be colored by that ping. Flutter echoes occur when the sound source is between two exactly parallel and smooth surfaces, and the resulting sound modification effect is to produce a ping as the sound is trapped between the parallel surfaces, being reflected back and forth. Flutter echoes should be avoided if at all possible; continue moving around the space testing with handclaps to find a spot without them (they are usually localized over quite a small region of sound source positions).

When performing on a stage, it is important to be aware that the stage is acoustically like a small room linked to a larger room (the auditorium). This is particularly the case when there is a proscenium arch at the join and the sound energy levels arriving in the auditorium can be considerably lower than those in the stage area. Once again the best advice is to listen to the space and get to know its sound. Ask a colleague to stand and speak or sing from the performer's position on stage and move around both the stage and the auditorium to obtain a listener's ear view of the sound output from the space in different positions as well as a comparison between the sound energy levels on- and off-stage.

In some situations, background noise can be a serious issue. This is less likely in a space designed for public performances, such as theaters, lecture halls, and opera houses, but relatively high levels of noise are sometimes present in other spaces where one is asked to perform vocally, such as classrooms, church halls, or back rooms in pubs, clubs, or hotels. These are the unwanted sound sources as illustrated in Figure 3–2, and they can either be internal or external. Noise can obscure or mask other sounds, and in the context of speech or singing, certain sounds of the language can be lost to the listener depending on the nature of the noise, thereby compromising intelligibility.

The sounds that are affected by noise will depend on the nature of the noise itself. Noise that is hisslike contains high frequency components and it will tend to mask sounds with high frequency components (see Chapter 2) such as fricatives. Buzzlike noises, such as the drone of electric motors in ventilation or air conditioning systems, will tend to mask frequency components in the formant frequency range, potentially masking vowels. Traffic or aircraft noise tends to cover a wide frequency range and therefore has the potential to mask all sounds. Short bursts of noise, perhaps from hammering, can affect the perception of plosive bursts.

One way to combat noise is to speak or sing at a higher level, but this is not an ideal solution as it places undue strain on the performer's voice. It is worth listening to a space to appreciate the nature of the unwanted noise—it is rare to find a space with no unwanted noise! Sometimes the source of the noise can be turned off. Perhaps a heater or ventilation fan could be turned off at the start

of the session, turned on again in the interval, and off again for the second half. If the noise is external, perhaps from traffic or aircraft, then it is worth ensuring that windows are closed properly, noting that if they are required to be open for ventilation, they could be closed at the start of the session, opened in the interval, and closed again for the second half.

The sources of some noises may not be identifiable or may not be locally controllable. Usually a noise will be worse in some parts of the space, and if this can be identified by careful listening, there may be scope for changing the layout of the room to avoid any noise hot spots. One other form of noise that cannot be checked beforehand is that produced by the audience. Have a strategy thought out for dealing with audience noise such as chatter, mobile telephones, or late arrivals. Raising the performance level to combat such noises rarely works since the noise levels will often be raised also, and this does have an acoustic energy cost that could result in vocal problems if it becomes a persistent habit. One method that can work very well is silence—a silent pause accompanied with a stare towards the noise source can be most effective and it has no acoustic energy cost!

Vocal Performance Considerations

A key element of vocal performance, whether singing or speaking, is communication of the text. The acoustic effect of the space can serve to destroy all endeavors to annunciate consonants carefully, because the overall sound reaching the ears of listeners who are well beyond the critical distance (see Sound Modification by a Space) is "blurred" or "mushy," so that the acoustic detail that is important for distinguishing between individual sounds is lost, with the result that the message becomes very difficult to understand.

Figure 3–8 illustrates this effect in terms of how reverberation serves to extend individual sounds acoustically in time, and the potential there is for them to overlap. The figure shows the average pressure level of the two consonants in the word *beat* (SAMPA: /bit/), for the situations where /b/ is louder than /t/ and /t/ is louder than /b/. The reverberation decay is indicated for a short and a long reverberation time (RT_{60}); this is how the energy of each consonant would decay. Notice that the later consonant is the one that could be obscured or masked by the reverberant decay of the first, and that this is the case for the long RT_{60} shown. When the later consonant /t/ is louder than the first /b/, complete masking cannot occur. Deliberate articulation of final consonants is therefore particularly important in highly reverberant spaces to combat this effect and ensure that the listener is provided with the acoustic cues required to enable all sounds to be heard clearly and identified. In addition, shortening individual syllables or sounds and lengthening the gaps between them will make space to allow the reverberation to die away before the next acoustic event.

One of the less well understood and compensated for acoustic issues facing anyone using his voice regularly in front of a group of people is the extent to which he can hear himself adequately. This is vitally important, because our vocal output is continuously being monitored by our brain, known as *self-*
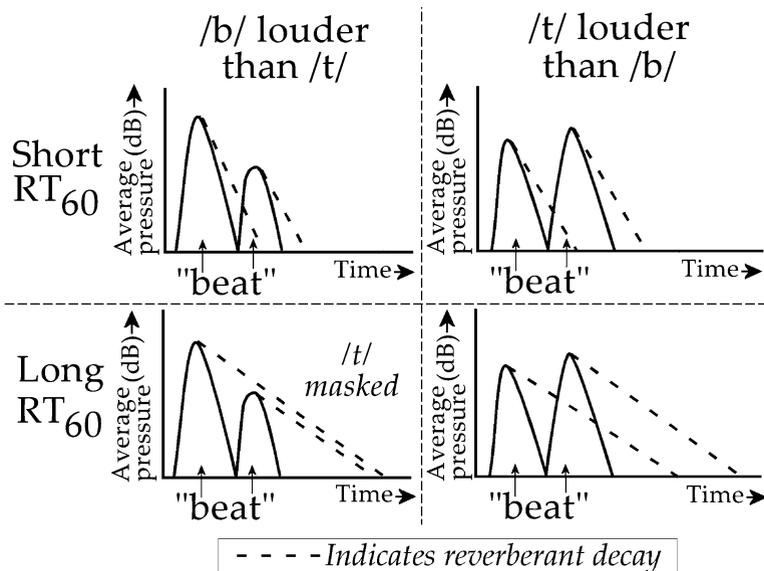
Figure 3–8. *Illustration of the potential masking effect of a short and a long reverberation time (RT$_{60}$) on the second consonant /t/ in beat (SAMPA: /bit/) by the first consonant /b/, depending on their relative levels.*

monitoring. If the acoustic energy level reaching our ears is insufficient, the brain insists that we speak louder so that it can carry out its monitoring job effectively. This is what triggers a tendency to speak or sing louder than is comfortable (and often louder than is actually necessary as far as the audience is concerned), and therein is a common route to vocal strain or more serious vocal problems. This is particularly an issue for those such as teachers or lecturers who use their voices over long periods. Teachers are the largest group presenting for medical attention relating to vocal issues such as strain, laryngitis, or hoarseness. So often, the cause is simply an inability to self-monitor adequately.

Increasing the level of the direct sound that is fed back to the ears of the performer can help to sort this issue out, and this can be achieved in a relatively straightforward manner. Starting with an awareness of the issue, creative possibilities can be thought through to improve matters through the use of acoustically reflecting surfaces placed close to the performer to provide *foldback*. Such reflectors are usually placed behind the performer, so that they are not only of benefit to the performer, but they also enhance the overall sound level that reaches the audience.

Specially made reflectors do exist, but there are ways of moving towards improving foldback using existing materials or apparatus. Any hard surface, such as a white board or blackboard, will serve as an acoustic reflector to provide foldback, and often these are on wheels and can be readily moved. Acoustic screens are becoming increasingly available for use in orchestras and music groups to shield the ears of players

sitting in front of loud instruments (e.g., percussion and brass) following renewed concerns about overall sound levels in public spaces. In some areas (in the United Kingdom this last occurred in April 2006), the maximum levels allowed have been lowered, prompting the provision of acoustic screens on health and safety grounds. Where such screens, which are usually made of thick acrylic and mounted on wheels, are available, try them out as reflectors to enhance what the performer is hearing.

When a reflector is set up to provide foldback to the performer, it is important to ensure that the delay involved is not too great. Delay will be incurred because the path travelled by the acoustic wave from the performer to the reflector and back to the performer's ears takes time, which can be determined from the velocity of sound (usually taken as 344 meters per second—see Chapter 1). A delay of no more than about 25 ms will typically not interfere with the benefits provided by the foldback, and in that time sound can travel 8.6 m. (The velocity of sound is 344 meters per second, so in 25 ms [or 0.25 s], sound will travel a distance equal to the time multiplied by the velocity, which is $344 \times 0.25 = 8.6$ m.) Remembering that the sound has to get to the reflector and back, this means that a foldback reflector should be placed within 4.3 m (half of 8.6 m) of the performer. In a small room, there is likely to be at least one wall within 4.3 m of the performer that will serve to provide foldback. However, in a large hall there probably will not be a surface that is close enough to act as a foldback reflector, and that is when it is worth setting up a local reflector.

When working on stage with scenery, it is worth being aware of the potential benefit of improving foldback, especially when performing a major speech, song, or aria. If a speaker is suffering unanticipated vocal strain or a singer is experiencing unexpected tuning difficulties, this could be due to a lack of an appropriate foldback level. With the director's agreement (or instruction), they could perhaps be positioned within 4.3 m of a piece of scenery that could provide acoustic foldback, or perhaps the scenery could be appropriately added to or rearranged.

The importance of foldback has been known to professional pop groups for a long time. Vocal performers are provided with local loudspeakers to provide an appropriate mix of the overall sound output as foldback, which includes their own vocal output. These loudspeakers are usually placed on the stage, so as not to be visually intrusive, with their loudspeakers facing upwards at 45 degrees towards the ears of the performer to localize the sound. Due to the shape of the loudspeaker cases, they are often referred to as *wedges*. Alternatively, foldback can be provided directly into the ears of vocal performers via bud earphones. This is known as *in ear monitoring* and it is especially useful for performers who move around a lot. For vocalists, it is vital that the foldback balance of relative levels of the different instruments and their own vocal output is appropriate. Good working relationships between whoever is in charge, the sound team, and the vocalists are an important part of vocal care in such situations.

When performing with acoustic instrument accompaniment, it is important to ensure that the accompaniment can be heard comfortably to maintain intonation and a proper balance for both performers and audience. It is always

worth moving around to find the best position acoustically, and then to consider how that position might be modified to provide best visual presentation to the audience. Often, just a small repositioning can make a huge difference to the overall acoustic coherence of a performance while also serving to provide less chance of problems for vocalists.

SUMMARY

This chapter has introduced sound transmission in a space from the sound source to the microphone or ear of the listener. The concepts of direct sound, early sound, and reverberant sound enable the nature of the sound output from a space to be better understood in terms of how this output sound changes for different sound source or listener positions in a space, and the possibilities that exist for practical and inexpensive acoustic modifications that might be made to a space.

When performing as a vocalist in a space, it is essential to remember that the acoustics of the space itself are a part of the performance. The acoustics of the space should be used to best advantage. Performance spaces are often well set up acoustically, and often there is little choice as to the performance position. Everyday work spaces are rarely set up acoustically, and there are a number of things that can be considered that can make it easier both for listeners to comprehend the message and for the performer's voice. Any use of the voice in any job is a performance. Considering it as such while taking advantage of the advice given herein has the potential not only to improve the quality of vocal life,

but also to gain esteem from those observing the performance. You only have one voice; look after it and use it well.

FURTHER READING

Acoustics of Spaces

Beranek, L. L. (1954). *Acoustics*. New York: McGraw-Hill.

Howard, D. M., & Angus, J. A. S. (2001). Room acoustics: How they affect vocal production and perception. In P. H. Dejonckere (Ed.), *Occupational voice—Care and cure* (pp. 29–46). The Hague, Netherlands: Kugler.

Howard, D. M., & Angus, J. A. S. (2006). *Acoustics and psychoacoustics* (3rd ed.). Oxford, UK: Focal Press.

Modifying the Acoustics of a Space

Everest, F. A. (1984). *The master handbook of acoustics*. TAB Books.

Everest, F. A. (1989). *Acoustic techniques for home and studio* (2nd ed.). TAB Books.

Howard, D. M., & Angus, J. A. S. (2006). *Acoustics and psychoacoustics* (3rd ed.) Oxford, UK: Focal Press.

Performing to Best Advantage in a Space

Howard, D. M., & Angus, J. A. S. (2001). Room acoustics: How they affect vocal production and perception. In P. H. Dejonckere (Ed.), *Occupational voice—Care and cure* (pp. 29–46). The Hague, Netherlands: Kugler.

Potter, J. (2001). *The Cambridge companion to singing.* Cambridge, UK: Cambridge University Press.

Smith, B., & Sataloff, R. T. (2000). *Choral pedagogy.* San Diego, CA: Singular.

been working hard for some time. Computer hard drives can also be sources of noise. In such a situation it is very worthwhile to try to source quiet equipment that will not cause such problems. Non-computer based stand-alone recorders or flash memory devices that have no mechanical parts might be good compromises. Alternatively, arrange additional absorbing material around the noisy equipment and place it carefully away from the vocal source in a corner of the chamber if possible. The other consideration with bringing in additional equipment is that it may be a possible cause of acoustic reflections (e.g., a laptop screen), and so careful positioning and the use of additional absorbing material should be considered.

The second problem to consider is the nature of the environment itself and its impact on the vocal performer. Listening to one's own voice in a completely reflection-free environment is a very unusual experience and can become tiring after prolonged exposure. Regular rest breaks in the session might be required, or alternatively, the performer should have a good headphone foldback mix with some additional reverberation added. If the latter approach is used, care should be taken to keep levels low to avoid headphone sound leaking into the microphone.

MULTIPLE VOCAL SOURCES

Introduction

Whereas the entire purpose of the previous section was to use microphone and recording techniques to remove the solo vocal source from the surrounding environment, this section is concerned with capturing the entire acoustic event—the sound of the voice source in the acoustic space within which it has been placed. Generally, this implies that there is no longer a solo performer but rather a number of vocalists that make up the specific sonic event. These vocalists may be spatially distinct and therefore considered as separate solo sources or, more commonly, be grouped together in an ensemble such as a choir. Stereo microphone techniques are also introduced as these methods are generally more appropriate for capturing the complete source plus environment combination rather than just the individual source. The one exception to this that was not included in the previous section is soloist recording in a performance venue without sound reinforcement. In this scenario, the desire is again to capture the sound of the complete event rather than the individual, and a stereo recording is the most appropriate solution. Hence, this special case is covered as part of what follows.

Stereo Recording

Under normal listening conditions we use our two ears, separated by our head, to locate the direction from which a sound source is originating. We rely on a number of auditory mechanisms to help us determine this source direction and they are summarized as follows.

Interaural Time Difference (ITD)

Depending on direction, the sound source will arrive at one ear before the other, resulting in a very small amount of difference in arrival time between left and right ears. This only works for low frequencies.

Interaural Level Difference (ILD)

Depending on direction, the sound source will be louder in the ear that is oriented more closely to the source, resulting in level difference between left and right ears of up to 20 dB SPL at some frequencies. This does not work at low frequencies, as the sound wave will diffract around the head such that the level difference between the two ears is negligible.

Pinnae Cues

ITD and ILD only give enough information to locate a source in two dimensions. The actual shape of the outer ear imparts a direction dependent frequency characteristic on the incident sound that helps to resolve front-back and up-down differences.

Head Movement

When attempting to work out the direction of a sound source, very slight head movements act to constantly change the relative ITD, ILD, and pinnae cues so that a listener can more easily and quickly determine source location. Primarily this acts to minimize ITD and ILD values to a point that they are essentially zero. This implies that the sound source is either directly in front of or behind the listener, and pinnae cues—and of course sight—can help to determine the final source direction.

With knowledge of these directional properties of the ear, particularly those relating to ITD and ILD, it is possible to fool the ear into perceiving a directional effect through just a pair of speakers or headphones—what we typically refer to as *stereophonic* or *stereo* audio presentation. There are two main ways of recording stereo sound images for presentation over a pair of loudspeakers—*coincident stereo* and *spaced stereo* microphone techniques.

Coincident Stereo Microphone Recording

This technique uses a pair of identical directional (not omnidirectional) microphones, each connected to a separate audio channel. To create robust, stable sound images suitable for stereo presentation, it is considered important to minimize time differences between left and right channels, and so the two microphones must be placed as close together as is physically possible—hence the term coincident stereo (also known as, XY, crossed pair, or normal stereo). The result is that sound sources are captured with differing levels between the two channels. This is due to the directional characteristics of the microphones used and the fact that signal amplitudes will vary in direct relation to the physical angle between the microphone pair and the sound source. Hence, this technique works well at presenting realistic and natural stereo sound because it is based on how the ILD works with the auditory system. The normal method is to place the capsule of one microphone immediately above the other, so that they are coincident in the horizontal plane, which is the dimension from which sound image positions will be created.

Coincident stereo is actually based on simple amplitude panning, as implemented in the mixing desk channel pan control (see Chapter 4, Routing and Output section), where altering the relative level of a mono signal between two stereo speakers will cause the sound image to be perceived as moving between them. This same relative panning effect can be replicated for a sound source moving between two coincident figure-of-eight microphones at an angle of 90 degrees to each other. This arrangement of microphones is called a *Blumlein*

pair, after Alan Blumlein, who first experimented with these techniques in the 1930s, and the resultant polar pickup patterns are shown in Figure 5–4.

The Blumlein pair arrangement gives accurate stereo imaging between two loudspeakers of the original position of the acoustic source. Note that the stereo image will be reversed for sources to the rear. Ideally, the 90 degree angle should be maintained and all relevant sources should be within the *acceptance angle*; this is defined as the usable working area in front of the microphone as defined by their polar patterns, and as also used when considering single microphones, as already discussed in the section Microphone Directivity Patterns in Chapter 4. Acceptance angle will therefore act to restrict the range of possible source to

microphone distances that will result in good stereo playback. The acceptance angle for a Blumlein pair is 70 degrees.

If cardioid microphones rather than figure-of-eights are used, the angle between the capsules needs to be wider in order to produce the same relative level differences between microphones for a given source position. This angle is actually taken as the point at which the response drops by 3 dB relative to the on-axis position, and for cardioids is defined as 131 degrees, as shown in Figure 5–5. This gives a much wider acceptance angle of 130 degrees, meaning that the pair can be moved closer to the source.

In practice, 90 degrees (giving an acceptance angle of 170 degrees) is the commonly used angle of separation for a crossed pair of cardioids, as it is easiest to set up, although it is possible to change this angle over a small range to

Figure 5–4. *The directivity patterns of a crossed pair of figure-of-eight microphones, used for coincident stereo recording. The angle of separation is 90 degrees, and this is also known as a Blumlein pair. This arrangement will result in the same stereo imaging as a standard mixing desk channel pan pot.*
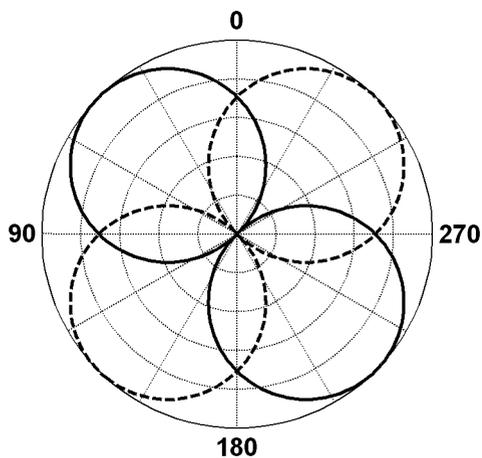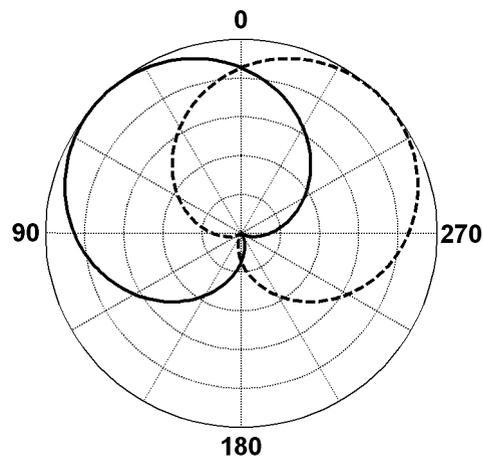

Figure 5–5. *The directivity patterns of a crossed pair of cardioid microphones, used for coincident stereo recording. The angle of separation for level differences equivalent to a Blumlein pair is 131 degrees.*

adjust the precise relationship between the physical sound source positions in front of the microphones and their perceived positions in the stereo image. Greater than 130 degrees will leave a "hole-in-the-middle" of the stereo field where there will be a noticeable drop in level and where central sound sources will fall outside the optimal polar pickup angle of each microphone. If the angle is smaller than 80 degrees, the acceptance angle becomes greater than 180 degrees, with the small focused overlap of the microphone directivity patterns meaning that lateral sound sources are considerably attenuated. A coincident pair based on hypercardioid microphones will be halfway between a cardioid and figure-of-eight, and hence the angle between the capsules should be 105 degrees.

One problem with the crossed pair techniques is that the center of the vocal source is off-axis from both microphones. This can lead to timbral coloration due to the less-than-ideal off-axis frequency response of the microphones. *Mid-side* (M-S) recording uses one microphone to capture the *middle* signal, which would be obtained if the outputs of a stereo crossed pair were added together. The other microphone captures the *side* signal, which would be obtained if the output of one microphone was subtracted from the other. The most common arrangement is to use a cardioid microphone facing forward (the mid microphone) together with a figure-of-eight microphone (the side microphone) facing sideways at 90 degrees, as shown in Figure 5–6. When these M-S signals are converted into normal left-right stereo, they produce an identical acceptance angle to conventional crossed cardioids.
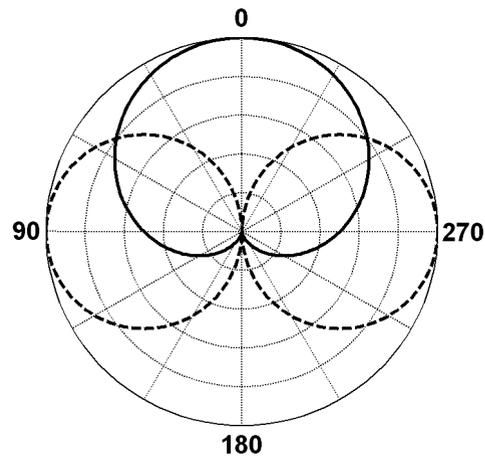
Figure 5–6. *The directivity patterns of a mid-side coincident microphone arrangement. The mid microphone is a cardioid that will be on-axis with the source (rather than off-axis with a crossed pair). The side microphone is a figure-of-eight. After conversion to normal stereo, this arrangement allows manipulation of the stereo image acceptance angle and to some extent the balance of direct to reflected sound.*

The two signals have to go through a conversion process before being auditioned on loudspeakers or headphones as in normal left-right stereo. The most useful aspect of the system for everyday recording tasks is that the acceptance angle and hence the perceived spread of sound sources across the stereo image can be controlled easily from the mixing desk or even after the recording. This will also to some extent allow control over the direct to lateral sound—giving some control over the amount of reverberation received relative to the direct sound. As the level of the mid microphone is increased relative to the side microphone, the useful acceptance angle and hence the perceived width of the stereo field is increased. As the level of

the mid microphone decreases the acceptance angle and relative stereo field also decreases. The M-S signals are converted to conventional stereo as follows:

1. Pan the M microphone to the center.
2. Split the S microphone to feed a pair of adjacent channels.
3. Pan the S channels hard left and right, phase reversing the right channel.
4. Listen with the monitoring switched to mono and balance the gains of the two S channels for minimal output.
5. Revert to stereo monitoring, and fade up the M channel.
6. Adjust the balance between the M and S signals for the desired image spread.

Note that it is also possible to use specific stereo microphones consisting of multiple microphone diaphragms in a single capsule, which makes the process of setting up for a stereo recording session significantly easier. In some stereo microphone designs the internal arrangement is fixed, in others it is variable and some control over the stereo image will be facilitated through external controls. Designs exist based on both the crossed pair technique and the M-S arrangement, with the latter giving somewhat greater flexibility for stereo field manipulation as part of the post-recording editing/mixing process.

Spaced Microphone Techniques

This method uses two (or more) identical but spaced microphones, each connected to a separate audio channel as before. However, with this technique, as the microphones are spatially separated, a sound will arrive at each capsule at a slightly different time according to the relative distance between it and the source. Hence, the spaced microphones effectively receive time-of-arrival information, and so this technique generates a stereo image based on timing differences (rather than level differences) and is comparable to how ITD works with the auditory system. However, the final stereo image, when presented over loudspeakers, is less stable and robust when compared with a similar coincident recording. This is because the sound emanating from each loudspeaker will arrive at both left and right ears (rather than left speaker to left ear only, right speaker to right ear only), and there will be a slight time delay added to the additional signal received at the opposite ear due to the off-center positioning of each speaker, as shown in Figure 5–7. This additional set of ITDs imparted onto the ITDs of the original spaced microphone recording tends to lead to confusion in terms of where a source is perceived to originate from.

An additional disadvantage—although less critical in modern audio distribution—is that if the outputs from the spaced microphones are mixed together to produce a single mono signal, the timbre of the overall mix might be altered due to phase cancellation effects (compare with problematic reflections when miking a single source and how direct and reflected path can cause the same effects). The greater the number of combined microphones, the worse the effect is likely to be. The big advantage of spaced miking, however, is that this technique allows the use of omnidirectional microphones, as the relative level of the acoustic source and how it varies with direction is not critical. The implication
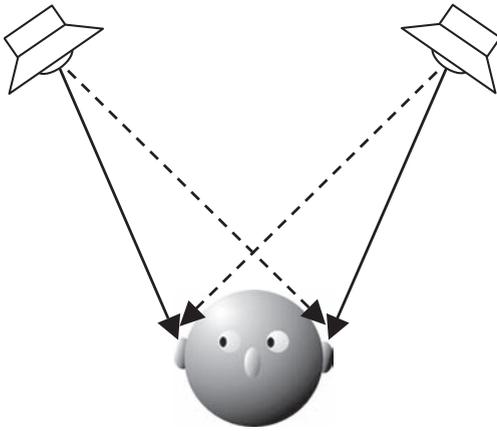
Figure 5–7. *When listening to stereo material using two loudspeakers, an additional set of interaural time difference auditory cues will be imparted onto the audio signal. This is due to the fact that the signal from each loudspeaker will arrive at both ears, and in each case the dotted line path is slightly longer than the solid line path, implying a time-of-arrival difference. This can confuse stereo information that is already dependent upon time differences captured from spatially separate microphones, leading to poor stereo imaging.*

of using such microphones is a significantly improved frequency response, particularly in the low end, and a more natural and transparent sound.

Although not necessarily accurate or stable in terms of stereo imaging, the final sound is normally perceived as having width and a certain amount of imaging information, and it usually sounds more spacious than a coincident recording. The recording might also suffer from a hole-in-the-middle if the microphones are too widely spaced apart. The simplest spaced microphone technique is to place an identical pair of omnidirectional microphones a distance apart in front of the sound source. A microphone spacing of between about a half and a

third of the width of the actual sound stage [the width of the source(s) that must be recorded] is a good place to start in terms of positioning. This could be improved through the use of additional directional microphones to alleviate any potential hole-in-the-middle effects. Other spaced techniques that use directional microphones are often called *near coincident* techniques because they combine the level difference recording characteristics of directional coincident microphones with spaced arrays. For instance, the *ORTF* method uses a pair of cardioid microphones with a separation angle of 110 degrees spaced about 17 cm apart from one another.

Multiple Soloists

Working with multiple soloists is based on an extension of the techniques introduced in the section, Single Vocal Sources, when recording the solo voice. As with the other applications discussed so far, the final listeners should be first considered when approaching the problem— what perspective will they have on the presented sound, and what is important for them to hear? If the multiple voices are supposed to be heard in the context of the space in which they are presented, it is probably best to consider them as an ensemble and to use a stereo recording technique to capture the sound of the overall event (see the section, Ensemble Recording). However, this will generally only apply to a limited set of possible applications, such as classical-style music performance, and there are a great many other situations where clarity of direct sound, separation, and control are more important factors to be considered. Examples might include studio recording of

multiple lead vocal lines, sound reinforcement of lead and backing vocals in amplified performance, interviews for radio or television broadcast, musical theater, and even applications such as teleconferencing. In each case, once the individual vocal sources are captured they will be subject to further manipulation, editing, or studio processing post-capture.

Sources and Location

In terms of location, studio work is generally the easiest to control as there should be enough space to separate the vocal performers, and this can be improved through the use of acoustic screens. Generally, the acoustics of a studio space will not influence the final recording too much, particularly if care is given to source and microphone positioning, and this will of course be helped further by using directional microphones. The main problem to consider will be spill from each performer bleeding into the other microphones, so achieving good sound source separation is important. However, problematic spill may well be masked when placed in the context of the wider music production and so might not be too serious a problem, although every effort should be made to minimize its effects. Working with multiple solo performers in an anechoic environment will offer many of the advantages of the studio in terms of control and clarity of direct sound. Care should be given to the possibility of problematic reflections, however, and there might not be as much space available in which to arrange the vocalists themselves. Studio based radio broadcasting will be a very similar situation given that the environment is optimized for the best sound for spoken voice, with good separation between announcers/

broadcasters, and little influence from the surrounding acoustics.

Just about every other example of working with multiple vocal performers will introduce problems related to the recording environment that will have to be considered and solved. Live sound reinforcement work is a simple extension of the solo vocalist case, although with every additional performer and microphone used will come the requirement for additional foldback monitoring and therefore another potential source of feedback. Every separate channel of monitoring should have a dedicated pre-fade auxiliary send from the main FOH mixing desk (if a separate monitoring desk is not being used) and most importantly its own graphic EQ for tuning out problematic frequencies. These vary with positioning and source-microphone distance, and so each microphone/monitor combination will require its own individual EQ settings. Also the more microphones that are used on stage, the more spill picked up from other off-axis sources. Therefore, non-critical microphone channels should be muted when not in use, or noise gates used to achieve the same effect automatically. Again, this will help to minimize possible feedback and generally make the mixing process easier. Obviously, directional microphones should be used at all times.

Other forms of broadcasting, particularly those based in television studios, will make use of lavalier/lapel/tie-clip microphones to ensure good capture of direct sound from each individual. Care should be taken in positioning the microphone well. The "tie" position works best, as it will center the microphone with the mouth, although it should not be so high that the chin shad-

ows the microphone, perhaps when the person's head moves to read written notes. Generally, such microphones also have a rolled off bass-response to take care of possible proximity effects if it has a cardioid family directivity pattern, and to compensate for a possible bass-rise from the resonance of the person's chest cavity. Similarly, the high end may be boosted slightly to compensate for the microphone being positioned off-axis. In general, once attached, the resulting sound should be checked for any major problems or variability as the talker goes about his regular business. This should include checking that the microphone sound/positioning is not affected by the subject's normal movements.

These types of microphones, usually mounted in a headset, are also used extensively in theater productions (especially musical theater) where the challenge is to get good sound from each actor while keeping the microphone and cables out of sight as much as possible. They are often only used for the principal performers, and so for capturing, for instance, the chorus or other on-stage sound, shotgun microphones are arranged around the edge of the stage to give good overall coverage. Graphic EQ might again have to be used to minimize feedback, although the levels involved are generally considerably less than amplified sound reinforcement and so are less of a problem. Another type of microphone used in this and similar vocal applications is the *boundary microphone*—sometimes also called a *float*. These microphones generally have a hemispherical (half omnidirectional) pickup pattern and are designed to be mounted on a flat surface (such as the front stage area or a wall). They operate on a different principle from the micro-

phones introduced so far (although they are generally a condenser type design) and use the acoustics of direct and reflected sound at a surface to give effective rejection of nondirect/reflected sound in favor of direct sound only (and hence will also help to minimize feedback problems). However, despite their good rejection of background sound, given their almost omnidirectional behavior, they should be used with care in terms of source pickup, applied gain, and possible feedback. Also, as they have to be physically placed on a wall or floor, they will be subject to possible mechanical vibrations from, for instance, footsteps.

In research applications, if multiple talkers or singers must be captured simultaneously, with clarity, separation, and accuracy being the important influencing factors, probably the only option available is to use headset microphones. If the recording location is large, with a controlled acoustic such that good separation can be achieved and reflections and spill minimized, there may be more options available, and the use of spaced omnidirectional microphones (favored for research applications) might be possible. However, this is generally unlikely given that the favored space for this approach would be a large anechoic chamber. Generally then, the headset method is perhaps the best way forward and has been reported as being successful in a number of studies. Note that additional microphones may be required to ensure that the headset microphones are calibrated correctly, as detailed in the section Recording for Research.

Where there is a requirement to capture multiple talkers and individual lavalier/headset microphones are not practical or possible, a simple one-to-one arrangement of spaced directional

microphones is probably the best option. This may be used for instance in round-table style conference presentations, or in multiple talker teleconferencing. In the latter application, if source separation is not critical a single (for mono) or multiply arranged (for stereo) boundary microphones placed appropriately will provide a good compromise. As with theater sound, care should be taken to avoid possible sources of mechanical vibrations that will be captured due to the microphone being directly placed on a local surface—typically a tabletop in this case—with drumming fingers or moving papers, pens, etc., being obvious sources.

Potential Problems

In the majority of these applications, the purpose of the sound system used is to ensure that the direct sound is captured in such a way that is it clear and separate from background sounds, particularly the other sources present in the local vicinity, so that control can be obtained over each individual source. However, as soon as more that one microphone is used—for even a single source—then the application verges on being a spaced stereo recording, even if single channel capture is the required outcome. This means that the related problems already discussed in this and the previous chapter are combined and must be considered and dealt with. These can be summarized as follows.

Omnidirectional vs. cardioid family. Omnis will give a flat frequency response and transparent sound at the expense of loss of separation and increased spill and background sound (including reverberation and reflections). Cardioid family microphones will improve separation and reject nondirect sound but poten-

tially alter the timbre of the captured vocal source.

Source separation and background noise. Priority is given to the capture of the individual direct voice source and rejection of background noise to facilitate maximum control of individual sources. However, no microphone is ideal, and every microphone added to a particular scenario will capture every source present to some degree, reducing separation and ultimately compromising control.

Off-axis colorization. Nondirect sounds will be captured off-axis by other nearby microphones. The off-axis frequency response of a directional microphone is not at all ideal, leading to potential timbral colorization problems when multiple channels are summed or auditioned together—giving a total result that consists of optimal on-axis direct sound from the source microphone, together with colored, off-axis spill captured from another adjacent microphone.

Phase cancellation effects. Spaced microphones imply captured timing differences for a particular source. This may lead to phase cancellation based timbral colorization due to the delay between signals when summed or auditioned together and/or confused stereo imaging. This problem will be more pronounced if there are major reflections also present.

As a helpful guideline, wherever possible, aim to have the distance between spaced microphones at least three (and preferably closer to five) times that of the individual source-microphone distance. In summary, maintaining good separation is the key to achieving a good recording or good sound in these and

other similar multiple solo source scenarios. Time, experimentation, and positioning will all help in this regard and ultimately provide greater flexibility, control, and creative options in editing, mixing, and post-production stages.

Ensemble Recording

Ensemble recording here refers to the capture of vocalists performing in a particular venue. Source separation is not the key focus; rather, priority is given to the accurate and transparent capture of the whole sound event, to include all performers, their relative balance, blend, and spatial positioning, and the acoustics of the venue itself. The most common example of where these techniques are appropriate would be choral recording, although this may include much smaller ensembles such as a barbershop quartet or the performance of spoken word. Note, however, that with smaller ensembles, production or aesthetic preferences may dictate that they should be recorded as multiple individual sources, as discussed in the previous section, rather than as a natural whole ensemble. Sound reinforcement is less common, as ensembles are sized for the material and space in which they will be heard, although individual microphones may be used to help soloists as part of the wider group (for instance spoken narration plus choral performance), and the use of such *spot microphones* will be considered as part of this section. In this context, the ensemble considered will also include solo performers where individual close-up miking is less appropriate and the sound in its totality in combination with the space is the desired result. Ultimately, what is being attempted is

the capture and permanent recording of a particular sonic event so that the listener hears what the audience would have heard in the venue (often in the best seat) during the performance itself. Hence, stereo microphone techniques are the primary method of realizing this desire.

Source and Location

In general, the aim is to arrange the ensemble around the stereo microphone arrangement in such a way that they occupy the complete stereo image, or rather to place the microphones so that they capture the complete stereo sound stage. Considering a large choir, as might be found accompanying an orchestra, as the vocal ensemble source, coincident crossed cardioids might be best placed above and close to the conductor position in order to achieve the desired stereo image width. However, a Blumlein pair of crossed figure-of-eights would have to be positioned a long way down the venue, much further away from the choir, to achieve the same stereo width due to their narrower acceptance angle of 70 degrees. In choosing the polar patterns for the stereo microphone arrangement, the physical separation between sound sources and microphones is determined for a given stereo width and therefore the listener's perspective of the recording.

In this example the cardioids would give a very close-perspective sound, with little reverberation, due to both close positioning and directivity pattern influenced rear rejection, and a distorted choral balance favoring those singers closer to the front and more centrally positioned. The figure-of-eights would give a much more natural and balanced perspective to the choir, but would also

capture a great deal of the reverberant sound, due to both their positioning and their increased rear pickup, which might make the recording rather more distant than anticipated. A compromise solution might use crossed hypercardioid mikes at some midpoint between cardioid and figure-of-eight extremes, or a scattering of close spot microphones to reinforce the weaker sections of the choir.

Microphone Positioning

As a starting point, the most commonly used method and easiest to get good results while still allowing a degree of experimentation is the crossed pair. These should be good quality cardioid condenser microphones, positioned one above the other, angled at between 90 and 120 degrees, according to source width. As mentioned above, they will also help to cut down on the amount of reverberation captured in the recording, due to the null pickup point at their rear. In general, a good level of direct sound compared with the reverberant sound is required to ensure overall clarity. Placing the microphones in a typical audience seat location will usually result in too much reverb when auditioned over loudspeakers. Experiment with distance to achieve the best balance between closeness/clarity and liveness/reverberance. It helps when setting source-microphone distance to have an estimate of the critical distance/reverberation radius of the space where direct and reverberant sound for a particular source are theoretically balanced. Clearly, the microphones should not be placed beyond this distance. It obviously helps when experimenting with position in this way to be able to listen directly to the results from the microphones to aid

the decision-making process. If possible, stereo loudspeakers in a separate control room should be used, although realistically for most scenarios, good headphone (preferably enclosed) listening will be the main method of monitoring the microphone signals.

Once an optimum distance for the microphones has been decided the stereo imaging produced by the coincident pair should be considered. Monitor the performers over headphones and listen to make sure that what is heard agrees with what can be seen and heard in the actual venue. If the stereo spread is either too wide or too narrow, then the angle of separation (and hence the acceptance angle) can be adjusted. If the stereo image appears off-center, ensure that gain levels are equal for both microphone channels and that they are pointing in the appropriate direction. If possible, before the performers enter the venue, make a recording of someone walking across the front of the performance area from stage left to stage right and listen back to the result to make sure the recorded stereo imaging is in good agreement.

Always use stands that give good stability and allow the microphones to be raised to a good, high level. Raising or lowering the microphones in this manner can also achieve a good balance between any soloists and the accompanying ensemble if this is required, or the ensemble as a whole if it consists of many people. Use shock mounts wherever possible to minimize possible noise from vibrations, or place the stands on some rubber or sponge mats to decouple them from the floor of the venue (carpet tiles can also be useful). A good starting position in terms

of microphone distance and height is about 12 feet from the performers and about 12 feet above the floor.

Potential Problems

If it is not possible to achieve an optimum balance between direct and reflected sound or stereo width due to limitations of time in setting up or restrictions in terms of microphone positioning, a combination technique might have to be used. This consists of a main stereo pair together with individual spot microphones. This technique is also appropriate if additional minimal sound reinforcement is required for particular aspects of the ensemble or for soloists working with them in combination, or if an omnidirectional spaced pair is used as the main microphone arrangement resulting in a hole-in-the-middle stereo imaging problem. A spot microphone is basically a close-up microphone used in combination with a more distanced main stereo pair to reinforce or generally improve the overall balance of sound sources. There are three things to consider with this combination technique.

Image position. The main microphone pair will establish stereo image positions for each aspect of the ensemble, and the close-up spot microphones should not contradict this virtual sound stage. Hence, each spot microphone must be panned appropriately to match the main stereo mix. The best technique for setting the individual pan positions is to concentrate on the stereo image of a particular vocal part from the main stereo pair, then slowly fade up the corresponding spot microphone, paying particular attention to how the image moves in the stereo field as this happens. If the image

pulls to the right, fade the spot microphone down, adjust the pan control slightly more to the left, and try again. Repeat until the pan position of the spot microphone is in agreement with the main stereo pair.

Perspective. A microphone close to the vocal source will have a completely different perspective to one much further away. This contrast is usually undesirable as it will draw undue attention to the soloist in question. The relative balance between the direct sound from the spot microphone and the overall direct-plus-reverberation mix from the main pair is critical. If the signal from the spot microphone is too noticeable then it is too high in the overall mix.

Timing. Note that this is usually only a problem with very large recording venues. Consider again the recording of a large choir in a large venue where the main stereo microphones may be 50 feet away from the main ensemble. Sound travels at approximately one foot per millisecond (ms), and so the signal from the stereo pair will be delayed by about 50 ms relative to any close spot microphones. Therefore, to ensure that close-up and distant microphones are in agreement, the spot microphones must be delayed by the appropriate amount, usually at the editing or mixing stage. It is therefore important in this case to measure source-microphone distances, and if appropriate, the temperature of the venue, which will help to give a good estimate of the speed of sound. From this a time delay can be calculated, with final adjustments made by ear to suit (see Chapter 1, section Sound Transmission and Velocity, and some of the

suggested further reading for more information relating to how the speed of sound varies with temperature and how this calculation would be performed).

Finally, if the stereo pair used, for whatever reason, tends to favor the direct sound from the ensemble, resulting in a slightly too close perspective, some additional control over the reverberation can be facilitated through the use of two additional spaced "ambient" microphones. These microphones should be placed beyond the critical distance of the space, with generally a spaced arrangement giving better results than a similar coincident pair, and they are used to capture the more reverberant sound of the venue. Once in place, they should be balanced up with the more direct coincident pair to give more control over direct/reverberant perspective of the finished result. However, both this and spot miking methods do add to the complexity of the overall recording (a mixing desk or multitrack recording system will be required rather than a two-track direct to stereo device), and so it is usually best to experiment with the main stereo pair to get the best sound, balance, and perspective that is possible for a particular source/location/listener combination.

Background Vocals in Popular Music

Background vocals in studio recording work are sometimes dealt with somewhat differently from what has been considered so far and are therefore considered separately here. Note that background vocals in amplified sound reinforcement applications should be treated as individual soloists for clarity, separation,

control, and minimization of feedback. This may also apply in the studio according to production or aesthetic considerations. However, it is generally assumed that in studio work there will be adequate control over the acoustics of the actual recording space and so a little more flexibility is allowed.

General background vocals (sometimes called "gang" vocals) involve grouping a number of performers around a single microphone rather than miking individually. This helps to provide a uniformity of sound, will give a particular energy to the recorded music, and allows the vocalists to interact with one another for the sake of the overall performance. A cardioid microphone, preferably a condenser, can be used for this, but the limitations of the acceptance angle and the possibility of off-axis colorization should be considered. As a result, there should be no more than two or three vocalists grouped in an arc around the front of the microphone. If the acoustics of the studio environment allow it, an omnidirectional directivity pattern would be ideal, providing a more transparent sound and allowing many more vocalists to be arranged around the microphone in a circle at an equal distance. As with stereo recording, the source-microphone distance in either configuration will alter the overall sense of perspective of the ensemble when auditioned.

If further stereo control is required to further enhance the overall production, it is relatively simple to replace the cardioid/omni with a coincident pair. The next level of improved separation and control would be to record each vocalist individually and at the same time. In this situation hypercardioid microphones will help to focus in on each individual

performer, and acoustic screens should be used if available. As with any technique involving multiple spaced microphones, the possibility of spill and phase cancellation effects should be considered, but this will allow individual panning of each source while helping to maintain the group vocal feel. Of course, the final level of control would be to record each vocalist separately and deal with them as any other solo vocal performer.

SUMMARY

This chapter has considered a wide variety of vocal recording and sound reinforcement applications and how they should be approached to achieve the best results. The definition of what might be "best" varies from case to case but should similarly be defined on a case-by-case basis through due consideration of the nature of the sound source, the environment in which it is being recorded, and perhaps most importantly of all, the demands or reasoning behind the final listening experience. There are many different scenarios where a vocalist has to work with a microphone and associated audio system, and a distinction has been made between single or multiple vocal sources. The particular demands of recording for vocal research have also been considered. Single source work introduces the importance of capturing the direct sound from the source while minimizing spill, background noise, reverberation, and/or feedback, with the implication this has for clarity, separation, and control over the final result. Ensemble vocal work introduces the use of stereo recording techniques and the importance of recording a complete

sound event in such a way that what the final listeners hear is a true representation of what they would have heard had they sat in the best seat in the house during the event. It should also be clear that this is a somewhat artificial delineation, as there are some scenarios that fall into both camps—for instance, recording a solo singer in a good concert hall.

It is important to note that the contents of this chapter should be considered as guidelines, rather than rules. Generally, there are no hard and fast rules when it comes to recording—the ultimate deciding factor is that the final result should sound "good." Again, the definition of *good* is highly subjective and will also vary according to the particular recording task. Making quality recordings for research purposes is a particular example here, although objective measures can also be applied to some extent in this case to determine the final quality of the results obtained. Some of the most important points to take away from this chapter actually have nothing to do with microphones and audio systems at all, but have everything to do with being a good sound engineer:

■ Plan for the recording session as much as possible beforehand.
■ Anticipate potential problems and possible appropriate solutions.
■ Allow plenty of time to set up.
■ Know the audio system well to get the best from it.
■ Test all aspects of the system prior to the start of the session.
■ Approach problems in a methodical and logical manner and consider possible alternative solutions before deciding on a course of action.
■ Experiment to the get best results out of a particular setup or scenario.

■ Respect the performers/artists, communicate effectively with them, and consider their own opinions in order to achieve the best possible results.

And finally, at all times listen carefully to the audio material: develop your listening skills, learn to trust your ears, and make an honest evaluation as to the quality of the final results in a bid to make even better vocal recordings in the future.

FURTHER READING

Bartlett. B., & Bartlett, J. (2005). *Practical recording techniques: The step-by-step approach to professional studio recording* (4th ed.). Oxford, UK: Focal Press.

Eargle, J. (2004). *The microphone book: From mono to stereo to surround—A guide to microphone design and application* (2nd ed.). Oxford, UK: Focal Press.

Granqvist, S., & Svec, J. G. (2005, September*). Microphones and room acoustics and their influence on voice signals.* Paper presented at PEVoC 6, London Retrieved from http://www.speech.kth.se/~svante/Granqvist_Svec_PEVoC6_Microphones.ppt

Howard, D. M., & Angus, J. A. S. (2006). *Acoustics and psychoacoustics* (3rd ed.). Oxford, UK: Focal Press.

Hugonnet, C., & Walder, P. (1998). *Stereophonic sound recording, Theory and practice.* Chichester, UK: John Wiley and Sons.

Jers, H., & Ternström, S. (2004, June). *Intonation analysis of a multi-channel choir recording.* Paper presented at the Baltic-Nordic Acoustics Meeting, Mariehamn, Åland, Finland.

Rumsey, F., & McCormick, T. (2005). *Sound and recording—An introduction* (5th ed.). Oxford, UK: Focal Press.

Talbot-Smith, M. (2004). *Sound engineering explained* (2nd ed.). Oxford, UK: Focal Press.

Talbot-Smith, M. (2004). *Sound engineer's pocket book* (2nd ed.). Oxford, UK: Focal Press.