1

Introduction to Australian English speech production

1.1 Introduction

Speech is the primary medium of communication for almost all human social groups. It is a highly complex skill that involves precise coordination of the structures of the upper and lower respiratory systems. The act of speaking is a uniquely human facility and although many animals have developed elaborate communication systems (e.g. songbirds), none has the ability to formulate novel audible (or, of course, written) messages of the type and complexity available to humans.

When we speak, we impart a vast array of information about ourselves in addition to the intended linguistic message. In this sense it is **extralinguistic** information. Extralinguistic features in language are considered to be indexical because they *indicate* to the listener something about the individual. Since speech is produced by coordinated movements of the lungs, the larynx and the structures of the mouth, including the lips, tongue and jaw, it also contains characteristics specific to the speaker that are determined by an individual's physiology. These physiological effects are referred to as 'biological indexical' features and include clues to our size, sex, age and health – aspects that are not under our voluntary control yet are fundamental individual speaker characteristics that allow our voices to be identified as uniquely ours.

Speech also has the potential to reveal something of our emotional or attitudinal state. Signs of emotion such as excitement or sadness, or of attitudes like boredom or enthusiasm, can be reflected in the way we speak, mainly through patterns of pitch and loudness, our rate of speaking and our voice quality. These features carry 'psychological indexical' information. Additional characteristics of speech can also give listeners clues to our social group membership such as the regional, socioeconomic, educational or gendered groups that we either belong to or aspire to. These are the 'social indexical' features (see Foulkes, Scobbie & Watt 2010). The distinctive speech patterns used by different social groups are referred to as accents. Everybody speaks with an accent, the characteristics of which are a consequence of the contact we've had with different people at different times in our lives. Accent also expresses the various identities that we bring to our social interactions. Humans are social beings and we are hard-wired to adopt symbolic behaviours that are similar to those of the people with whom we interact. Speech is therefore an important expression of group membership and it plays a vital function as a symbolic representation of who we are. Our accent reflects our social and cultural history, including aspects of our heritage language (e.g. speaking English with a French accent) and our regional affiliation/upbringing (e.g. speaking Australian English with an Adelaide accent or a Melbourne accent). Positive and negative opinions about accents are a response to the social meanings associated with an accent and not at all to the linguistic integrity of the accent itself. Each accent is equally capable of meeting the complex needs of its users.

During childhood and adolescence, identities are in flux and it is similar-age peers who are particularly influential in shaping accents (e.g. Kerswill & Williams 2005). Children and adolescents must learn to navigate linguistically and socially meaningful aspects of language to develop language behaviours that reflect and display their identities and social groups. Their accents are highly malleable but become established in the late teens and early twenties when a more stable sense of self and identity has developed, but they can continue to change across the lifespan (Harrington 2006), particularly if social/cultural circumstances or individual aspirations alter. It is important to acknowledge that within all communities accents change over time and the description of the speech sounds of a particular accent simply provides a snapshot of time, place and context.

Synchronic variation (i.e. variation in an accent that occurs at one particular time in history) is present in all accents. This variation may be a reflection of social conditions and determined by the age, gender, region of upbringing, social group affiliation, or language background of its various members. Such variation also relates to situational effects such as stylistic variability based on levels of formality and the context of any interaction. **Diachronic variation** (i.e. variation in an accent that occurs over time) is a

reflection of linguistic change. Diachronic variation has its roots in syn-chronic variation because such variability within a speech community can be the catalyst for both long- and short-term change.

Listeners to our spoken language are faced with diverse and complex signals about who we are and the message we convey, whether deliberate or not. All spoken social interaction involves so much more than simply sending and receiving linguistic messages, and includes the complex process of recognising, responding to and negotiating the interaction between the extralinguistic signals that accompany any interpersonal exchange.

In this book we will focus on **Australian English** (**AusE**), one of the core global dialects of the English language. We define **dialect** as a variety of a language that has its own specific and unique characteristics of vocabulary (the words and their meanings), **syntax** (the grammatical conventions that govern the creation of words, phrases and sentences and how they are combined to form meaningful utterances) and **phonology** (the structure and function of the sound system, including how sounds can be combined to form syllables, words and phrases, and how patterns of prominence and **intonation** influence the spoken output). **Accent** specifically relates to the phonological characteristics of the dialect. As the focus of this book is the speech patterns of AusE, we will concentrate on aspects of phonology and therefore the AusE accent.

Accents vary according to:

- **Phonemic system:** This relates to the categories of contrastive sounds (phonemes) that are used to make words. AusE has the same system of sound contrasts (phonemic system) as Standard Southern British English (**SSBE**) and New Zealand English (**NZE**), but differs from many American English (**AmE**) dialects. For instance, AusE, SSBE and NZE share a contrast between the vowel sounds in words like *balm* and *bomb* whereas most AmE accents don't differentiate these two vowels. These represent phonemic differences between the accents.
- **Phonetics:** The specific detailed characteristics of the individual speech sounds and how the sounds vary in different contexts are considered to be phonetic effects. For example, the sound represented by the letter *l* in AusE has a number of different variants that are selected according to the context in which they occur. For instance, the *l* sound at the beginning of the word *leaf* is different from the *l* sound at the end of the word *feel*. If you say these words carefully you should be able to hear and feel the difference between the two types of *l*. This is a phonetic difference to be discussed in more detail later in the book. In contrast to AusE, the *l* sounds at the beginning and end of words in Irish English are not very different from each other. In some accent varieties, such as for many speakers from South-East London and from South Australia (Horvath & Horvath 2002), the *l* sound at the end of a word

such as in *bell* or *bottle* may be more like a vowel, a little like the vowel in *pull*. These differences between various accent types represent phonetic differences between the varieties.

- **Phonotactics:** Restrictions on how sounds combine and how they are **distributed** in the construction of words are the phonotactic effects in language. For instance, AusE does not have *r* sounds before consonants or pauses in words like *card* and *car*, but Scottish English does. This is because AusE is a **non-rhotic** variety, where *r* sounds only occur before vowel sounds as in *red* and *borrow*. In **rhotic** dialects, such as Scottish English and many varieties of AmE, *r* sounds occur in all positions without the condition that they must occur before vowels.

- **Lexical features:** Sometimes words vary in how they are pronounced between different varieties of a language. These word-specific effects are referred to as lexical effects. For example, AusE speakers can use the vowel in the word *cat* or the vowel in the word *cart* when producing a word like *dance*. This difference varies regionally. The typical form in NZE, or AusE spoken in South Australia, contains the vowel sound in the word *cart* (Horvath & Horvath 2001).

- **Prosody:** The patterns of intonation, rhythm and prominence in language are referred to as prosody. For example, the stress pattern of the word *finance* varies across accents of English with either the first or the second syllable receiving prominence. This is an example of a lexical stress difference (i.e. the stress patterns that occur within individual words). If the first syllable of *finance* has prominence, it is said to be the stressed syllable in the word. Prosody also encompasses features that extend across entire utterances, including the shape of the pitch contour (how the pitch of the voice rises and falls over the duration of the utterance) and its relationship to patterns of prominence. We can think of the pitch pattern of an utterance as its tune. The characteristics of tunes vary across different dialects. Many AusE speakers, particularly when engaged in a narrative, use an intonation pattern sometimes referred to as Australian Questioning Intonation, **high rising tune (HRT)** or more recently as **uptalk** (Fletcher, Grabe & Warren 2005, McGregor & Palethorpe 2008, Warren 2016). HRT is a tune that occurs when statements are produced with rising pitch. Tunes of this type also occur in New Zealand, Belfast and Glasgow varieties of English.

Many of these phonemic, phonetic, phonotactic, lexical and prosodic aspects of language will be discussed throughout this book. Individual characteristics and details of the production of the vowel and consonant sounds will be given in Chapter 2. In Chapter 3, information about syllable structure, word stress and sentence stress will be described along with details of the prosodic

organisation of speech. In Chapters 4 and 5 broad (phonemic) and narrow (phonetic) transcription, the two different types of transcription that can be used to auditorily describe the sounds in speech production, will be explained. In these transcription chapters, an extensive set of exercises is available to assist you in developing competence in the different transcription techniques. All dialects in all languages display accent variation and AusE is no exception to this. AusE varies across speakers of different ages and genders, and from different regions and social/cultural/heritage-language groups, but it is beyond the scope of this book to provide explicit detail of this variation. For convenience we will therefore focus on **Standard (Mainstream) AusE** as the dominant dialect in the community. Standard AusE in this sense simply refers to the accent type used by the majority of Australians. It does not refer to any notion of correctness or prestige, and is a term adopted to differentiate the dominant variety from others in the community such as Australian Aboriginal English and Ethnocultural varieties such as Lebanese, Vietnamese or Greek heritage Australian Englishes, to name just a few. As you read through the material, you may like to consider how your particular accent corresponds to, or differs from, the typical accent type being discussed.

1.2 Phoneme and allophone

There are two main forms of **transcription** that can be used to document speech sounds: **broad (phonemic)** and **narrow (phonetic)** transcription. Broad transcription is the simplest form and uses a restricted set of symbols – those that represent phonemes. A **phoneme** is an abstract linguistic construct that exists in the mind of a speaker or listener. Consider the speech sounds in the following six words: *pan*, *nap*, *span*, *spring*, *naps* and *napkin*. Speakers of English agree that all these words contain *p* sounds but might not be aware that the *p* sounds are actually all a little different from each other. We think of them as the same (i.e. we categorise them all as *p*) but this *p* category only exists in the mind. It is an abstraction from all of the different variants of *p* that can be produced by speakers of the language. This abstract representation is a phoneme, and in this example the phoneme is /p/. It is not an actual sound, but represents a class of sounds. By convention phonemes are indicated with **slant brackets** (forward slashes).

Phonemes are important in the creation of words. For example, speakers of English know that the two words *pat* and *pad* have different meanings and that the final sound is responsible for the difference. *Pat* and *pad* represent a **minimal pair**. These two words contain the same sounds in the same sequence except for a single element, which in this case is the final sound. This minimal pair provides evidence that *t* and *d* represent two separate

phonemes in the language: /t/ and /d/. The difference is therefore a contrastive one. Selecting /t/ over /d/ changes the meaning of the word. We can follow this line of reasoning to identify all of the contrastive sounds in any given language. In English, the words *pack*, *pan* and *pal* show that the sounds at the ends of these words, *k*, *n* and *l*, also contrast with /t/ (*pat*) and /d/ (*pad*) – (*pack*) /k/, (*pan*) /n/, (*pal*) /l/ – and therefore form part of the phonemic inventory of English. Each language has a finite set of phonemes that can be used in various combinations to create meaning. For example, the speech sounds that we think of as *p*, *a* and *t* can be combined together to make the words *pat*, *tap* and *apt*.

The actual spoken variants of a phoneme in all of the different environments in which it occurs are known as 'phones'. Phones are entities that exist in the spoken output rather than just in the mind of the speaker. Those phones that are specifically conditioned by phonetic context are called **allophones**. For example, the different variants of the phoneme /p/ that occur in *pan* and *span* are context-dependent allophones of that phoneme. If you observe these two sounds carefully, you will notice that the /p/ in *pan* and the /p/ in *span* are actually quite different from each other. In phonetic terms the /p/ in *pan* is a voiceless aspirated bilabial stop (one that allows an explosive puff of air to be expelled from the lips) and the /p/ in *span* is a voiceless unaspirated bilabial stop (one that doesn't have the same amount of explosive air release). You can feel the difference in the degree of air expelled for each type of /p/ by placing your hand in front of your lips while you say the two words. These two *p* sounds are context-dependent allophones because one occurs at the beginning of the word (without any other consonants preceding it) and the other occurs after the /s/. Nevertheless, they remain members of the single phoneme /p/. As the differences between the different types of /p/ are not very important for speakers of English, we tend not to notice them. If you used the same type of /p/ from *pan* in *span*, the meaning of *span* would not be changed but it might sound a little odd.

Another example of a context-dependent allophonic difference can be found in the words *pat* and *pad*. You might be able to hear that the vowel is shorter in the word *pat* than in the word *pad* – a difference that we are not usually aware of. Such allophonic differences help to give the speech its natural sound and native-speaker quality but they are not the main cues used to differentiate the words in the mind of the speaker/listener.

The selection of allophones is dependent on many factors, including the rules/constraints or processes that determine how speech sounds in various phonetic contexts are produced in the language. The language-specific processes necessary to ensure the selection of appropriate allophones are 'known' to each speaker of the language, albeit subconsciously. The allophones are therefore predictable. A speaker of English 'knows' which allophone of /p/ to select when they say the word *pan* or *span* and they can

generalise to other words like *pin* and *spin*, or *pun* and *spun*. Foreign accent is often the result of a speaker using the processes specific to their native language to produce the speech sounds of a non-native language.

1.3 Broad (phonemic) and narrow (phonetic) transcription

If we want to convey information about speech production in written form it is necessary to use a standard set of symbols. It is not possible to simply use the letters of the alphabet for transcribing the sounds of language because there are only 26 letters but over 40 vowel and consonant sounds in English, and you will be very aware that English spelling conventions are not always transparent with regard to pronunciation. Notice, for example, that in AusE the words *her*, *bird*, *hurt*, *earth*, *worm* and *journal* all contain a common vowel sound but the vowel is represented six different ways in the spelling. Conversely, the words *dough*, *through*, *rough*, *cough*, *fought* and *drought* all contain the same sequence of letters (*ough*), but each word uses a different vowel sound. Consider words like *walk*, which contains four letters but three sounds (/woːk/); *knee*, which contains four letters but two sounds (/niː/); and *music*, which contains five letters but six sounds (/mjʉːzɪk/). These examples demonstrate that we can't rely on spelling to help us decompose a word into its speech sounds. In fact once we have learned to read and spell, our interpretation of speech patterns can be coloured by our knowledge of spelling conventions. Preliterate children tend to interpret the sounds in words phonemically and this can be seen in their early attempts at spelling (e.g. 'kwin' for *queen* and 'crusht' for *crushed*). The child seems to be interpreting the phonemes correctly but is not yet familiar with the letter-to-sound correspondences. Transcription of speech sounds requires that we divorce ourselves from the spelling conventions that are so powerful in biasing our interpretation of how words are pronounced. It might surprise you that the word *ink* does not actually contain an *n* sound at all. The sound represented by *n* in *ink* is the same as the one at the end of the word *sing*. This sound is symbolised by the letter sequence *ng* in *sing* (/sɪŋ/) but *n* in *ink* (/ɪŋk/).

It is important to use a standard system of symbols in order to facilitate written communication about speech sounds. Since phonemes are not actual sounds but instead exist as abstract linguistic constructs in the mind, they can be represented by any arbitrary system of symbols. The most widely accepted system for this purpose is the International Phonetic Alphabet (**IPA**). The IPA is the standard system used for representing the sounds of the world's languages. It is regularly reviewed and maintained by the International Phonetic Association (https://www.internationalphoneticassociation.org).

The *Handbook of the International Phonetic Association* (1999) outlines the theoretical assumptions behind how speech sounds are represented phonetically. The main idea is that each symbol represents a different speech sound and therefore a single symbol can be used for the same sound in any language that uses such a sound.

A true narrow (phonetic) transcription can ideally represent the phones of any spoken utterance along with all the details necessary for a speaker to reproduce the utterance regardless of the language in which it was spoken. In practice, though, we may want to instead represent the phonemes of a language and not the phones. Once the phoneme inventory of a particular language is known it is possible to select symbols from the IPA to represent those phonemes. The IPA does not dictate how phonemes should be analysed or transcribed for any particular language but it does equip those working in linguistics with the tools to make informed decisions and choices about the best selection of symbols to represent the phoneme inventory. The phonemic symbol set, once decided upon, can then be used for broad (phonemic) transcription.

When linguists are developing a phonemic description of a language or dialect, they are at liberty to select a subset of symbols from the IPA to represent the phonemes of the language. It is usual to represent phonemes using the phonetic symbol associated with the most common or widely distributed allophone of a particular phoneme. When an IPA symbol is used to represent a phone or allophone, it has a very different meaning from when the same symbol is used to represent a phoneme. The symbol for the phone/allophone is a closer characterisation of the actual sound, whereas the symbol for the phoneme reflects the group of phones that comprise the phoneme. For this reason it is necessary to indicate explicitly the type of transcription that has been made so the reader can correctly interpret the transcription. Broad (phonemic) transcription must be enclosed in forward slashes / … / and narrow (phonetic) transcription must be enclosed in square brackets […]. This bracketing convention immediately signals to the reader the type of transcription intended. The interpretation of the symbols is different in each case. As an example, take the symbol /k/, which represents the English phoneme that occurs in words like *kid*, *skid*, *scream* and *sick*. When the /k/ symbol is indicated with forward slashes, the reader interprets this as a phoneme and understands that its actual production can only be ascertained by listening to the sound and examining the characteristics of the allophone produced. The phoneme /k/ has a very large number of different realisations (phones) in speech, such as a voiceless aspirated, unaspirated or unreleased velar oral stop, a palatal stop, an uvular stop or a fricative variant, to name a sample of its realisations. Chapter 2 provides details of these descriptions. However, if the [k] is indicated with square brackets, the reader can be quite confident that the symbol more accurately

reflects production of the individual sound. The symbol [k] embodies a particular set of articulatory features and uniquely represents the voiceless unaspirated velar oral stop. The IPA chart reproduced at the back of the book shows this symbol in the column for the velar sounds in the top row.

Another important consideration is that narrow (phonetic) transcription is language-independent and therefore you don't need to know the linguistic processes relevant to any particular language for its interpretation. Broad (phonemic) transcription, on the other hand, is language-dependent and requires not only knowledge of the symbol–phoneme relationship but also knowledge of how a phoneme is realised in context in the specific language under investigation. The following examples illustrate this language-dependent relationship in two different languages.

In English, the word *pan* is transcribed phonemically as /pæn/. The initial sound is one of the allophones of the phoneme /p/. Speakers of English 'know' that /p/ in this position (i.e. as the single consonant at the beginning of a stressed syllable) is realised as the allophone [pʰ], the voiceless aspirated bilabial stop. If the reader of the transcription does not know the 'rules' for allophone selection in English – for instance, if they don't speak English as their native language – they may select an allophone that is not appropriate for the native-like production of the English word (such as the unaspirated sound that would occur in French). This would lead to a non-native or foreign-accented production.

Our second example relates to Fijian language. The town of Labasa is located in the Fijian Islands on the northern island of Vanua Levu. A Fijian broad transcription of the town's name, Labasa, may be similar to /labasa/. Without knowledge of Fijian phonology, a non-native speaker may have difficulty interpreting this broad transcription correctly. This is because in Fijian language, consonants like /b/ are prenasalised, which means that the /b/ in *Labasa* is pronounced with a preceding nasal sound like /m/, similar to the combination of *m* and *b*. The correct pronunciation of the town's name sounds like *Lambasa*. A Fijian-speaking reader of the broad transcription would be able to interpret the broad transcription correctly armed with intimate knowledge of Fijian language phonology and the processes required for selecting the correct allophones. However, an English reader who does not know Fijian phonology would be unable to interpret this broad transcription because English does not share the Fijian prenasalisation rule. You will notice that the spelling of the Fijian word follows the phonemic convention of just representing the phonemes and not the allophones. If this word were to be transcribed using narrow transcription conventions, anyone with knowledge of the IPA could reproduce the spoken word correctly because the phonetic transcription would contain the [mb] sequence (among other things). This specific detail would be required for accurate interpretation.

It is very important that you have a clear understanding of the difference between phonemes and phones and the related broad and narrow transcription representations. These concepts are fundamental to phonetic science and are prerequisite for understanding and using the material contained in the following chapters.

In summary, broad transcription is useful as a way of indicating which phonemes occur in different words and therefore in identifying the contrastive elements in words. The broad transcription doesn't contain detail about how each sound is produced in different contexts. It relies on the reader to supply that information when interpreting the transcription. It is used in dictionaries and as the basis for alphabetic writing systems where exact details of pronunciation may be a hindrance. Narrow transcription, on the other hand, is a system where many specific details of segmental pronunciation are expressed. Here the transcriber and the reader do not need knowledge of the phonological processes of the language in order to interpret the transcription. They do, however, require an intimate knowledge of the IPA and the ability to equate sounds of varying quality with the appropriate phonetic symbols.

1.4 Australian English transcription conventions

English speech sounds vary across dialects, and consequently the transcription of their phonemes can also vary. There is not a single phonemic system that can be considered most applicable for the general transcription of all English varieties combined. For this reason a number of phonemic transcription systems exist for the different varieties of English. There are two main systems used for transcribing AusE; these differ from one another mainly in their vowel symbols. The traditional system based on Mitchell (1946) and Mitchell and Delbridge (1965a) (henceforth MD) is employed by the *Macquarie Dictionary* (2013) and is derived from a transcription system originally devised for British English. This system indicates the congruence between the phonemic systems of SSBE and AusE. The revised system of Harrington, Cox and Evans (1997) (henceforth **HCE**) uses symbols that are reported to more accurately reflect the AusE variety. The revised set of symbols from HCE will be used as the basic system throughout this text but solutions to all broad transcription exercises in Chapter 4 will be given in both systems for the convenience of those who either want to learn both systems or who would prefer to select the system most applicable to their needs. An evaluation of the two systems for AusE transcription is provided in Chapter 6.