Contents

Foreword

Speech sounds have intrigued me since I first read about phonetics while I was a high school student. I did not expect then that I would pursue a career in which the sounds of speech would have a central and continuing place. I still marvel over the way in which thoughts and emotions are sent from one person to another through the invisible channel of sound. Having taught phonetics and speech science, I am familiar with the challenge of cataloging and describing the sounds of speech. Although we call them "speech sounds" they are not only sounds (auditory events), because they also can be described as articulatory actions that correlate with a certain auditory impression. Speech movements, like auditory events, are largely invisible without the aid of specialized technology. To understand phonetics, we must make visible those things that are usually invisible. We want to visualize the acoustic pattern that underlies the perception of speech along with the articulatory pattern that underlies the acoustic one. Only in this way do we come to a reasonably complete understanding of phonetics. So we come to rely on the tools of the speech scientist or phonetician, because these tools enable us to visualize the acoustic and articulatory domains of phonetics.

However, the array of available tools has not been brought to bear on the subject of speech sounds, at least not until the appearance of McLeod and Singh's ambitious and painstaking text, which documents speech sounds with the methods of cinematography, ultrasound, electropalatography, and spectrography. The rich descriptions afforded by these tools enable the student of phonetics or speech science to grasp the essence of speech sounds—how they are molded in the vocal tract and how they are expressed as acoustic patterns. The success of this text is proof of the adage that the whole is greater than the sum of its parts. Certainly, the sum of the parts is important enough, but a synthesis of the parts gives us the whole, in this case, the manifold nature of a speech sound. McLeod and Singh have marshaled phonetic tools and knowledge to create an authoritative yet friendly text that guides the reader through the phonetic landscape of the sounds of English. *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech* offers a comprehensive and lucid examination of the elements of speech.

With this book, speech sounds come into view (literally) as articulatory and acoustic images. This new text is a resource that is unrivaled among phonetics or speech science books, and it sets a standard for the use of technology to reveal the essential nature of the sounds that are the substance of speech communication.

Ray D. Kent, Ph.D.
Professor Emeritus
University of Wisconsin-Madison

Chapter 1

Visualizing Speech Sounds

OVERVIEW

As an atlas provides maps of the world demonstrating topography and landmarks, *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech* provides images (or maps) of the mouth for the production of speech sounds. These maps have been created using cinematography, ultrasound, electropalatography, and spectrography. Interpretation of each of the types of maps is described in depth in Part 1 of this chapter. An understanding of phonetic transcription and the International Phonetic Alphabet will assist users' understanding of the nomenclature used to describe vowels and consonants in this book. The International Phonetic Alphabet is provided in the front material of *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech*.

The production of speech sounds occurs by movement of air through the human respiratory and speech apparatus (Figure 1–1). The major anatomic structures involved in speech production are the tongue, lips, teeth, hard palate, velum (soft palate), larynx (vocal folds), nasal cavity, mandible (jaw), and lungs. Different configurations of these structures result in the production of discrete consonants and vowels. In the English language there are 24 consonants and at least 19 vowels and diphthongs. Each consonant can be differentiated according to their place of production, manner of production, and whether they are voiced or voiceless. Each vowel can be differentiated accord-

ing to the advancement and height of the tongue, and tenseness and roundedness of the lips. Part 2 of this chapter describes these features of consonants and vowels in depth. Readers who have prior knowledge of phonetics may wish to skip Part 2 and continue to Chapter 2. Each subsequent chapter profiles a different English consonant or vowel.

Speech-language pathologists (SLPs) frequently assess and provide intervention to "correct" the production of speech sounds. Inaccurate productions of speech sounds often are difficult to alter in therapy. This may be due to an absence of objective knowledge regarding the tongue positioning involved in the production of speech sounds (Gibbon, 1999b). Much of the information used within speech-language pathology clinics regarding typical production of speech sounds is based on perceptual analysis techniques such as impressionistic phonetic transcription. Weismer (1980, p. 50) indicates "if we are committed to a phonological analysis of speech sound errors; however, an analysis that relies solely on auditory skills is unacceptable." The description of speech can be enhanced using instrumental techniques such as photographs, filmstrips, spectrography, electropalatography (EPG), ultrasound, electromagnetic articulography (EMA), x-ray, and medical resonance imaging (MRI) (Ball et al., 2001; Hardcastle & Gibbon, 1997; Stone, 2005). These instrumental measures provide objective and detailed, real-time information about speech production. The next part of this chapter enables the interpretation of images created by the different instrumental techniques.
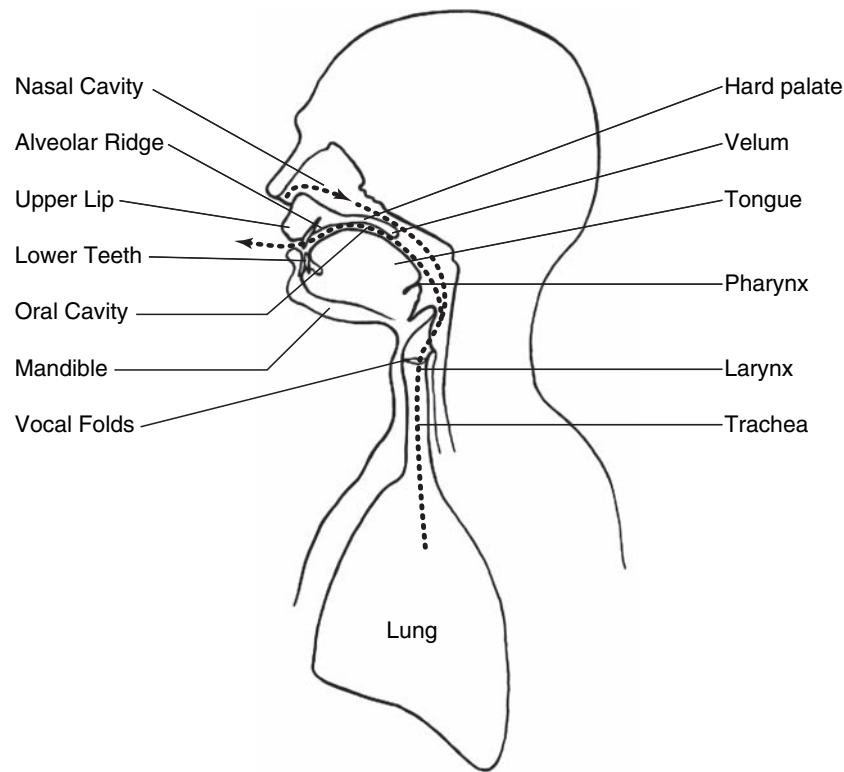
Figure 1–1. Human respiratory and speech apparatus (*Source:* Singh, S. and Singh, K. [2006]. *Phonetics: Principles and Practice* 9 [3rd ed.]. San Diego, CA: Plural. Figure 4–3c, p. 48).

PART 1: INTERPRETATION OF IMAGES OF CONSONANTS AND VOWELS

Each instrumental technique provides different insights into the production of consonants and vowels (see Table 1–1 for an overview). It can be seen from Table 1–1 that in order to comprehensively map speech sounds, a variety of imaging techniques are necessary. Images of speech sounds can either be static or dynamic. A static image is one distinctive image that is frozen in time and captures the most significant aspects of the consonant or vowel being studied. However, speech is a continuous stream of sound created by continuous movement of the oromusculature. As a result, it is not enough to segment speech into discrete units. Dynamic images, which demonstrate continuous production, are essential for understanding the contextual coarticulatory influences of consonants and vowels on one another. Thus, dynamic images typically consist of a series of frames extending over the production of an entire sound, word, or phrase. The purpose of *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech* is to make available the direct viewing of the phenomenon of speech production to facilitate a thorough understanding of the differences and similarities in the articulation of consonants and vowels produced under varying contextual influences.

The presentation of the total picture of phoneme production in context is contained within Chapters 2 through 36. With certain limitations, each chapter is a self-contained, complete story of a sound of speech as it is being produced. There are 24 chapters depicting the English consonants, 10 chapters depicting vowels, and one chapter for the diphthongs. Analytic comments are presented in each chapter. These comments focus on dimensions of speech production and speech acoustics that the authors consider important. The readers (viewers) are advised to examine them in detail, first one at a time and then comparatively. Chapter 37

Table 1–1. Features of Consonants and Vowels Viewed Using Different Technologies

	Feature	Photograph	Filmstrip	Schematic diagram	Ultrasound	Electropalatograph (EPG)	Spectrogram	Waveform	X-ray	Magnetic Resonance Imaging (MRI)
Consonants	Voicing			*			*	*		
	Advancement				*	*			*	*
	Labiality	*	*	*						
	Sonorancy						*	*		
	Continuancy						*	(*)		
	Sibilancy						*			
	Nasality			*			*			
Vowels	Advancement				*	*			*	*
	Height				*				*	*
	Tenseness	*	*							
	Rounding	*	*							
	Voicing						*	*		

assists in comparing images by placing similar imaging techniques side-by-side in tables of consonants and vowels. Chapter 38 provides detailed information about the creation of the static and dynamic images of speech production in this book.

INTERPRETING STATIC IMAGES OF CONSONANTS AND VOWELS

In Chapters 2 through 36, static images are presented to depict speech sounds in isolation. Each chapter contains at least 4 different static representations of a given speech sound: (1) photograph, (2) schematic drawing of tongue position, (3) ultrasound, and (4) electropalatography frame(s). Interpretation of each of these image types is described in turn.

Photographs

The photographs emphasize the front, or outside, of the mouth. They provide detail regarding the placement of the lips, teeth, jaw, and in some cases, the tongue at a single moment during the production of a consonant or vowel.

Schematic Line Drawings

Schematic line drawings are provided to visualize the overall function of the tongue—its elasticity, its flexibility, and its limits—all of which contribute to its systematic movements in the oral cavity. The line drawing accompanying each chapter emphasizes the position of the tongue during the production of a particular consonant or vowel. The voicing of a phoneme is indicated by the use of a plus or minus sign at the

approximate point of the larynx. In most images, the velopharyngeal port is closed, indicating oral airflow. However, in the line drawings for /m, n, ŋ/ the velopharyngeal port is open.

The schematic diagrams are extremely simplified. The tongue shapes are based on ultrasound images of the tongue during the production of each sound in connected speech. The schematic diagrams were created by superimposing these ultrasound images of the tongue onto a template of the nasoaoropharynx (see Chapter 38).

It is also important to note that the schematic diagrams represent only the midsagittal plane, that is, the midline of the body. Consequently, detail regarding the lateral margins of tongue movement is not included. This can lead to the misconception that the sides of the tongue are in similar configuration to the midline. This is particularly relevant in the production of alveolar sounds. For example, during the production of the following alveolar sounds, the midpoint of the tongue is raised to the alveolar ridge: /t, d, n, s, z, l/; However, comparison between the schematic line drawings and the electropalatography (EPG) images reveals that it is only the /l/ sound that has a similar shape along the lateral margins of the palate. For the other five sounds, the sides of the tongue are raised to rest near the teeth in order to provide lateral bracing.

Ultrasound

Ultrasound enables us to view the surface of the tongue during speech. To create an ultrasound a transducer is held below the chin and a wedge-shaped scan of sound waves emanates from the transducer. The sound waves travel through the body of the tongue until they reach the upper tongue surface. The white line on the ultrasound image is a reflection of the air above the surface of the tongue, or the tongue on the palatal surface. The thickness of this line is not relevant. Typically, approximately 1 cm of the tongue tip is not visible in the production of alveolar sounds due to the acoustic shadow of the jaw. Similarly, the tongue root may be obscured due to the acoustic shadow of the hyoid bone. Stone (2005) provides an extensive tutorial regarding recording and interpretation of ultrasound images.

Ultrasound images of the midsagittal plane of the tongue were taken during production of speech sounds and words. Figure 1–2 shows the tongue at rest, which can be used as a comparison with the other ultrasound images in *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech*. The bright white line is the surface of the tongue. The air shadow can be seen above this bright white line and the muscles and fatty tissue of the tongue can be seen below the white line. Chapter 38 describes the creation of the ultrasound
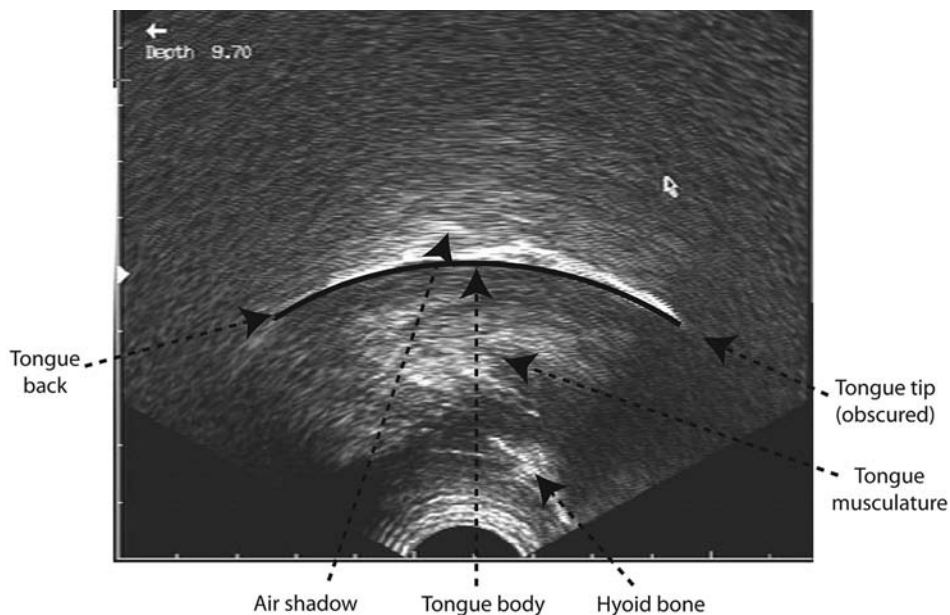

Figure 1–2. Ultrasound image of the tongue at rest.

images and the use of a purpose-built helmet to ensure stability of the ultrasound probe so that each image within this book can be compared with one another. In each image an arrow cursor has been included at approximately the level of the nose (top right quadrant), so that comparisons can be made.

The tongue tip is on the right for the ultrasound images. As for the schematic line drawings, it is also important to note that the ultrasound images only represent the midsagittal plane. It is important to compare the ultrasound images with the EPG images to determine the location and movement of the lateral margins of the tongue. Again, this is particularly important for the production of the alveolar sounds: /t, d, n, s, z, l/.

Figure 1–3 illustrates the relationship of the tongue to the palate in the ultrasound image of the production
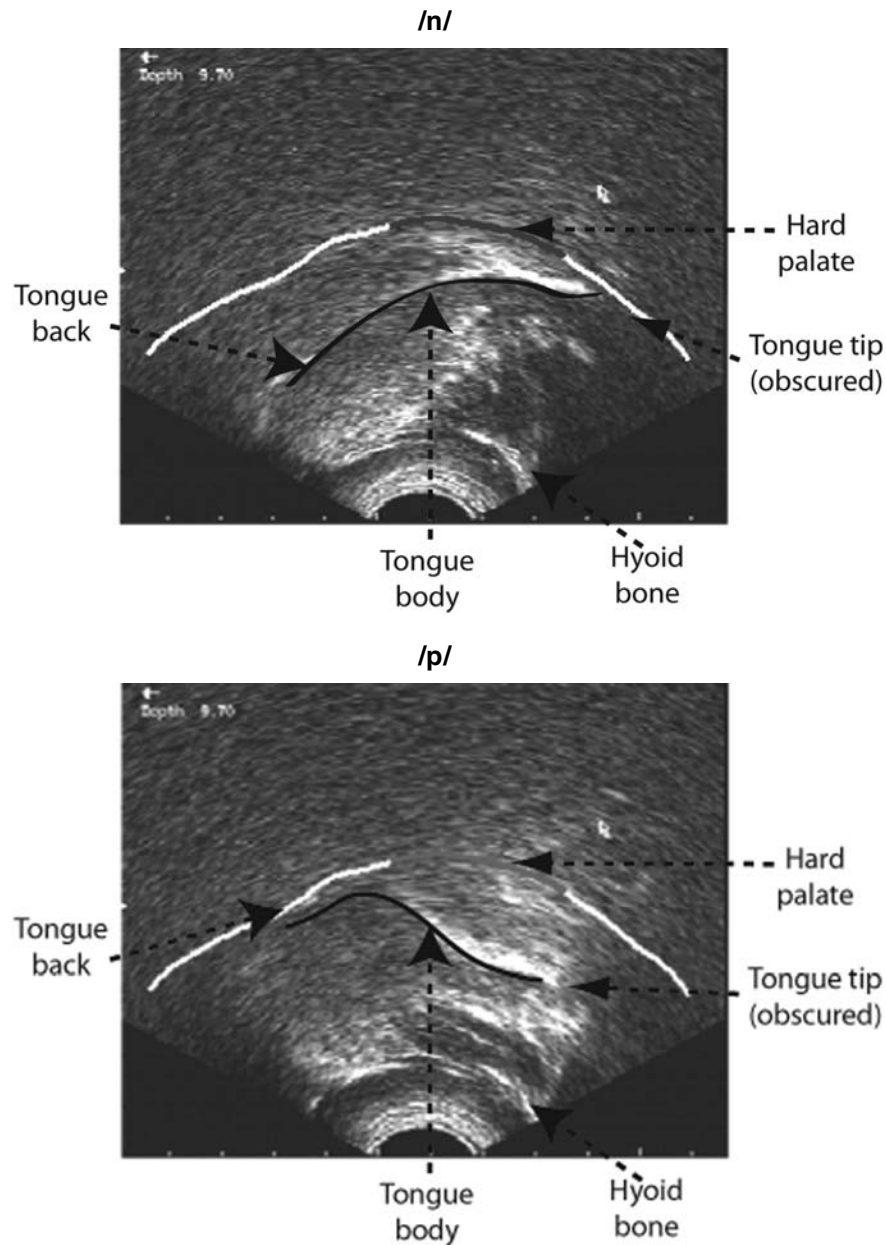
/n/

/p/

Figure 1–3. Ultrasound image of /n/ and /p/ with the palate drawn in the image. Note the bright white line representing the tongue. The tip of the tongue is at the right of the image. The tongue tip is up for /n/ and down for /p/.

of /n/ and /p/. At the beginning of the simultaneous ultrasound and electropalatograph recording, the speaker swallowed so that the imprint of the palate could be drawn; this was then superimposed on the subsequent image of the productions of /n/ and /p/. Below the bright white line of the tongue, there are diagonal lines reflecting muscles and fat within the tongue. There is also an air shadow above the tongue in this image.

Figure 1–4 compares simultaneous productions of ultrasound and electropalatographic images of the sentence "I see pop again." Readers can observe the difference in tongue position on the midsagittal ultrasound images and the coronal electropalatographic images. Readers also can compare the difference in tongue position for each sound within the sentence. For example, the tongue tip is up for production of

/s/ and /n/, but down for production of the vowels, /p/ and /g/.

Researchers of typical speech production have used the ultrasound to study aspects such as:

◆ tongue surface of English consonants and vowels (Stone, Faber, Raphael, & Shwaker, 1992; Stone & Lundberg, 1996)
◆ protrusion, grooving, and symmetry of the tongue during speech (Bressmann, Thind, Uy, Bollig, Gilbert, & Irish, 2005)
◆ production of schwa in /z/+ consonant sequences (Davidson, 2005)
◆ trough effect in coarticulatory sequences (Vazquez Alvarez, Hewlett, & Zharkova, 2004)
◆ swallowing (Chi-Fishman, 2005).

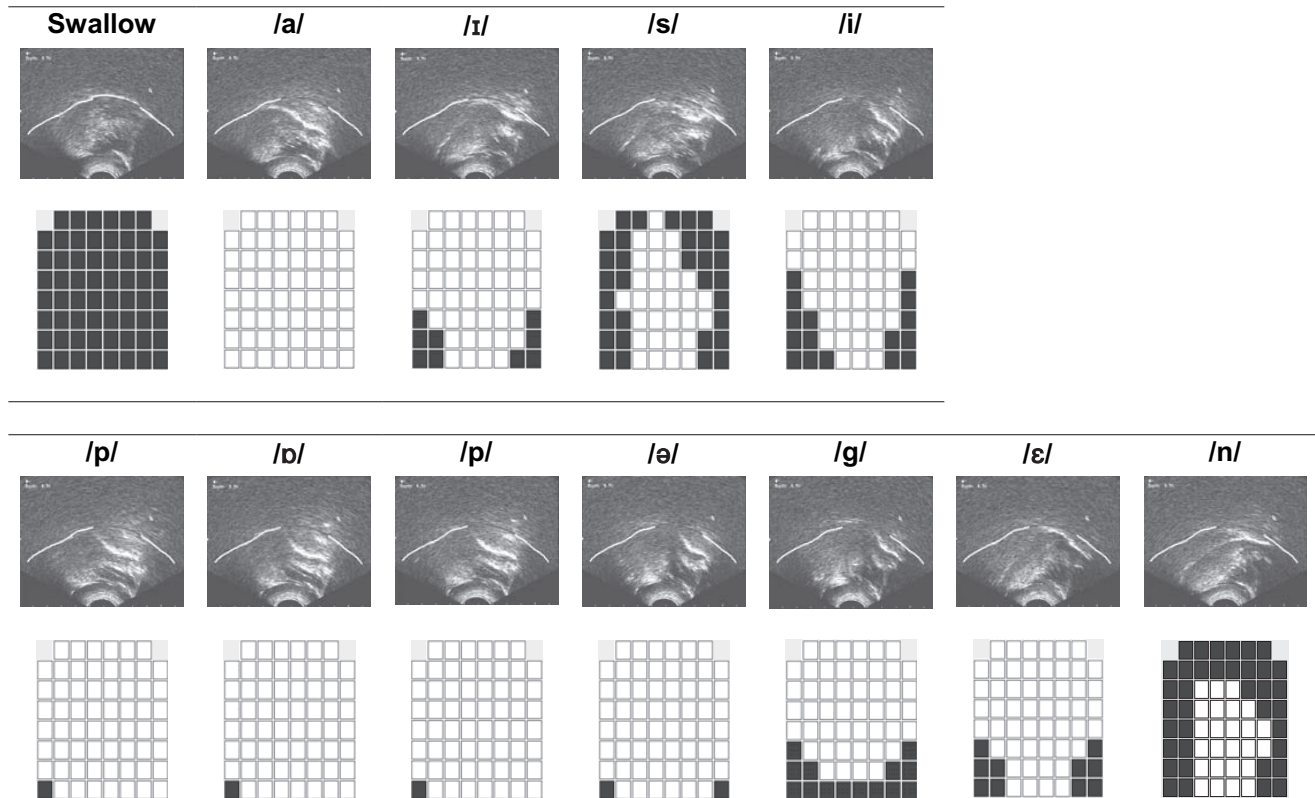**Figure 1–4.** Simultaneous ultrasound and electropalatographic images of the sounds in "I see pop again." *Note*: The right of the ultrasound image corresponds with the tip of the tongue. The palate trace has been drawn on the ultrasound images based on the swallow image to demonstrate the range of tongue movement. The top of the EPG image corresponds with the palate immediately behind the front teeth.

Researchers of people with speech impairment have used the ultrasound to study aspects such as:

◆ vowel production of adolescents with hearing impairment (Bacsfalvi, Bernhardt, & Gick, 2007)

◆ consonant production (/s, z, l, r/) of adolescents with hearing impairment (Bernhardt, Gick, Bacsfalvi, & Ashdown, 2003).

Electropalatography: Single Frames

The electropalatograph (EPG) records tongue contact with the palate during speech production, providing a printout of activated electrodes every 10 milliseconds.

The EPG has an artificial palate (Figure 1–5) that is individually molded to fit the roof of the mouth. The Reading/WINEPG used to create the images in this book has 62 electrodes that record tongue contact with the surface of the palate. The Reading/WIN palate is usually around 1.5 millimeters thick. The EPG provides a printout of the electrodes that are activated during speech. These are taken every 10 milliseconds. Figure 1–5 shows one frame. The black squares

represent the activated electrodes where tongue to palate contact has occurred. The white squares indicate that no contact has occurred. The areas of the palate can be separated into three zones: alveolar, palatal, and velar (Hardcastle & Gibbon, 1997). These are also presented in Figure 1–5. It can be seen that the top of the EPG frame corresponds with the alveolar region immediately behind the front teeth. The final row of squares corresponds with the juncture between the hard and soft palates. The maximum point of contact is often selected for analyzing the place of articulation for individual speech sounds as it provides a point of comparison (Hardcastle & Gibbon, 1997). This is selected by identifying the frame that has the highest number of contacted electrodes. An example of the maximum contact frame for /s/ is shown in Figure 1–5.

Electropalatography: Cumulative Frames

A cumulative EPG display is created by accumulating data from maximum contact frames for each consonant. Chapter 38 describes the study of eight speakers' productions of speech sounds that have been used in this book (McLeod, 2003). Figure 1–6 presents the
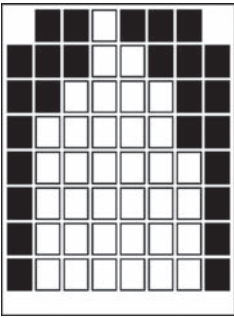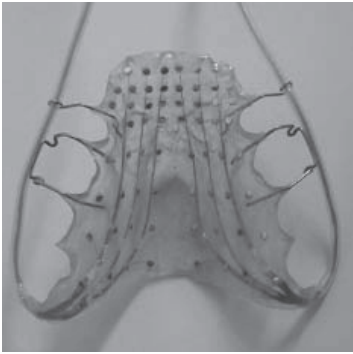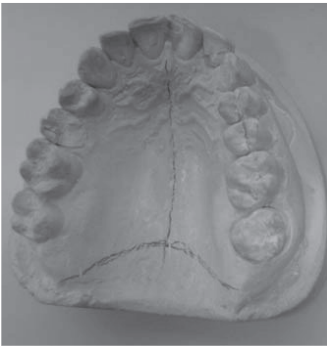
Row	EPG frame of the maximum point of contact for /s/	Region of palate	Hypothesized area of tongue	EPG palate	Dental cast with junction between hard and soft palate drawn
1		Alveolar	Anterior (tip)		
2					
3		(2 rows)			
4		Palatal	Blade		
5		(3 rows)			
6		Velar	Posterior		
7		(3 rows)			
8			(body)		

Figure 1–5. An example of an EPG frame for the production of /s/ juxtaposed with an EPG palate. The black squares represent the contacted electrodes. The zoning scheme is adapted from Gibbon (1999b). The EPG frame, EPG palate, and dental cast are from the speaker who produced the EPG images for *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech.*
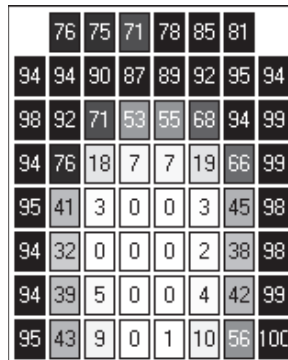
76	75	71	78	85	81		
94	94	90	87	89	92	95	94
98	92	71	53	55	68	94	99
94	76	18	7	7	19	66	99
95	41	3	0	0	3	45	98
94	32	0	0	0	2	38	98
94	39	5	0	0	4	42	99
95	43	9	0	1	10	56	100

Figure 1–6. Cumulative EPG display for eight English adults' productions of /t/.

cumulative display for /t/ for eight typical adults as shown in Chapter 4. The darker the shading on rectangle (representing the electrode) the greater proportion of tongue/palate contact. The number in each box indicates the percentage of occasions each electrode was activated by all the participants during their productions of /t/. For example, the bottom left-hand electrode was contacted 95% of the time by the participants during their productions of /t/.

Electropalatography: Single Frames Demonstrating Intra- and Interspeaker Variability

A range of EPG studies have been published that describe typical and impaired productions of vowels and consonants. A summary is provided in Tables 1–2 and 1–3. As can be seen, some sounds have received a large amount of attention. Others have received none. Gibbon and Paterson (2006) indicated that the most commonly treated sounds using the EPG were /s, z, t, d/ and this is evident in Table 1–3.

It was considered important to demonstrate variability in the images presented in research. Many of the following chapters include additional EPG images from others' research in order to demonstrate variability in the production of speech sounds across English accents, languages other than English, children, and people with speech impairment. Specifically, *Speech Sounds: A Pictorial Guide to Typical and Atypical*

Speech includes images of differences between studies of:

◆ adults speaking English,
◆ children speaking English,
◆ adults speaking languages other than English, and
◆ people with speech impairment.

The selected images have been redrawn with permission to simplify the complexity and individuality of the published images. A simplified EPG template was created. A black rectangle indicates that that electrode was contacted at least 67% of the time in the originally published EPG image. A gray rectangle indicates that that electrode was contacted less than 67% of the time. There are a few exceptions to this where the published studies have used different criteria (for example, Guizik and Harrington (2007) used a criterion of 50%); however, any departure from the 67% is mentioned in the text.

INTERPRETING DYNAMIC IMAGES OF CONSONANTS AND VOWELS

In Chapters 2 through 36, dynamic images are presented to depict speech sounds while they are being produced contextually. Each English consonant is presented in one or more contexts in a meaningful word (initial, within word, and/or final positions). Each of the vowels and diphthongs is presented within meaningful words to show the influence of consonants on vowel formants. For example, Chapter 2 depicts the English phoneme /p/ at the word-initial and word-final positions in the word "pop" as well as /p/ at the word-initial and within-word positions in the word "puppy." A careful study of these two words independently and simultaneously provides a better understanding of the phoneme /p/.

Each chapter contains at least 3 different dynamic representations of a given speech sound: (1) filmstrip, (2) sound spectrogram, and (3) a series of electropalatography frames. By carefully studying these three segments, the reader can develop an understanding of the correlation between articulation and acoustics.

Chapter 2

/p/

The consonant /p/ is a voiceless bilabial stop.

Place of articulation: bilabial

Advancement: front

Voicing: voiceless

Labiality: labial

Sonorancy: nonsonorant (obstruent)

Continuancy: noncontinuant (stop)

Sibilancy: nonsibilant

Nasality: nonnasal (oral)

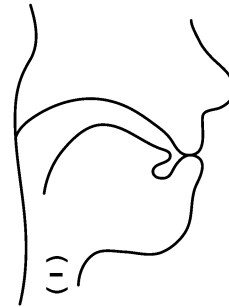## STATIC IMAGES OF THE ARTICULATORY CHARACTERISTICS OF /p/

The static images of /p/ are presented via a photograph, schematic diagram, ultrasound, and electropalatograph images (Figure 2–1).
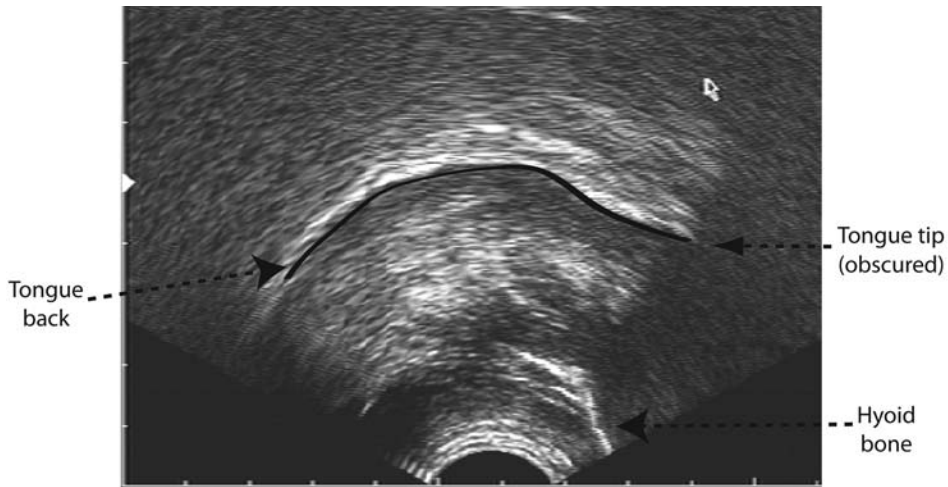
Photograph

In this photograph (Figure 2–1A), the upper and lower lips are approximated to create closure of the airstream to produce /p/. The front view of the lip approximation can be compared with the lateral (side) view of lip approximation in the schematic diagram.
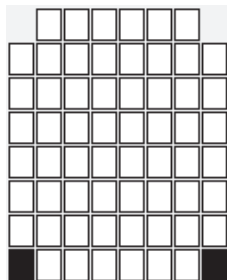
A. Photograph

B. Schematic diagram

C. Ultrasound

D. Electropalatograph (EPG) (single frame)

Figure 2–1. Static images of the articulatory characteristics of /p/.

Schematic Diagram

The involvement of both the upper and lower lips in the production of /p/ is shown by the drawing of lip approximation in Figure 2–1B. The tongue is not involved in the production of this labial consonant as shown by its location at a rest position. The symbol (–) has been chosen to indicate that the vocal folds do not vibrate systematically in the production of this speech sound.

Ultrasound

The bright white line on the ultrasound image of /p/ in Figure 2–1C shows the air above the tongue surface. The tongue tip is on the right and is toward the floor of the mouth. The back of the tongue is somewhat raised in the oral cavity. An air shadow can be seen above the tongue and diagonal muscle fibers can be seen below the surface of the tongue.

Electropalatograph (EPG)

The primary oral mechanisms involved in shaping the oromusculature to produce a /p/ are the lips and the soft palate (see Figures 2–1A and 2–1B). There is extremely limited tongue/palate contact for the EPG image of /p/ in Figure 2–1D. In this image the tongue touches the most posterior corners of the palate, as shown by the dark boxes. There is no contact with the central regions of the palate. In other productions of /p/, it is possible that there could be no tongue contact with the palate during production of /p/. Gibbon and Crampin (2002) stated that "simultaneous valving [that is, complete constriction] in the linguapalatal region throughout the period of lip closure is not a feature of normal speakers' productions of bilabials" (p. 41). Gibbon, Lee, and Yuen (2007) found that the extent of tongue/palate contact for the production of /p/ was significantly correlated with the extent of contact for the surrounding vowels.

DYNAMIC IMAGES OF THE ARTICULATORY AND ACOUSTIC CHARACTERISTICS OF /p/

In order to obtain a comprehensive view of the production of this consonant, the dynamic aspects of the production of /p/ are shown in a filmstrip, spectrogram, and EPG images in two words containing /p/ in different word positions (Figures 2–2 and 2–3).
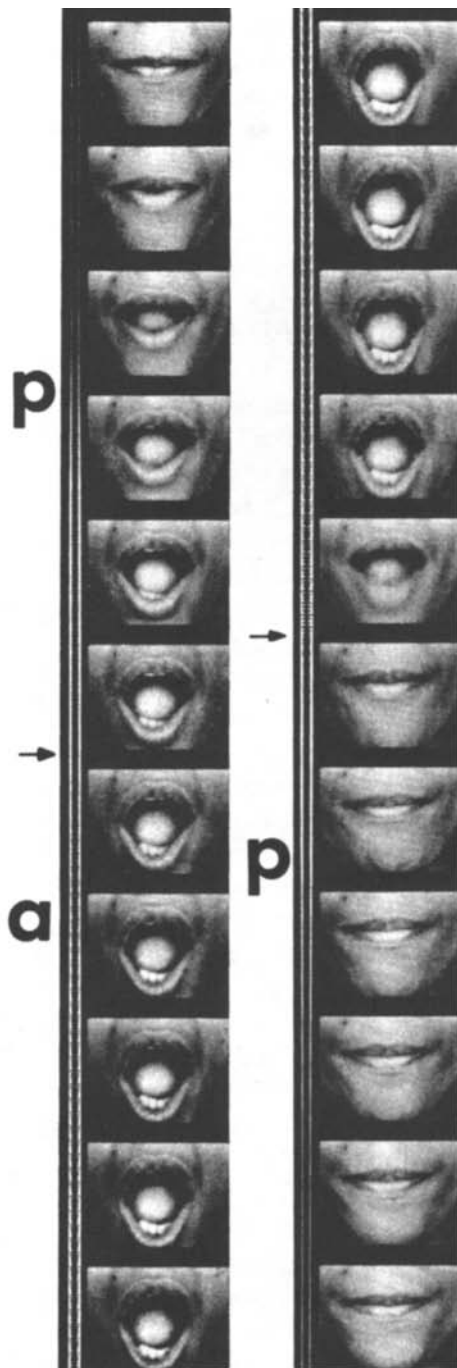
/p/ in Word-Initial and Word-Final Contexts: "pop"

Filmstrip: /pɑp/ "pop"

The phoneme /p/ is at the initial and final positions of the word /pɑp/ (Figure 2–2A). The complete closure of the lips can be seen in the first frame, a slight opening in the second frame, and a full opening pertaining to the target vowel /ɑ/ in the sixth frame. The subsequent 9 frames (7 through 15) sustain the lip position exclusively attributable to the vowel /ɑ/. The next 7 frames (16 through 22) gradually show the accomplishment of the final target /p/. Because final stops in English are sometimes unaspirated as in this case, no lip opening can be seen in the final frame. The arrows along the sides of these strips indicate the general location of the transition between adjacent phonemes. The first arrow marks the transition between the initial /p/ and the vowel /ɑ/. The second arrow marks the transition between /ɑ/ and the acoustic energy for the final /p/. The vowel /ɑ/ is clearly marked in the sound track by waveforms characterized by high amplitude and a repetitive pattern.

Sound Spectrogram: /pɑp/ "pop"

The phoneme /p/ is presented in both the initial and final positions in the word "pop" (Figure 2–2B). Lip involvement for /p/ is associated with energy mainly in the low-frequency region of the sound spectrogram.

A. Filmstrip
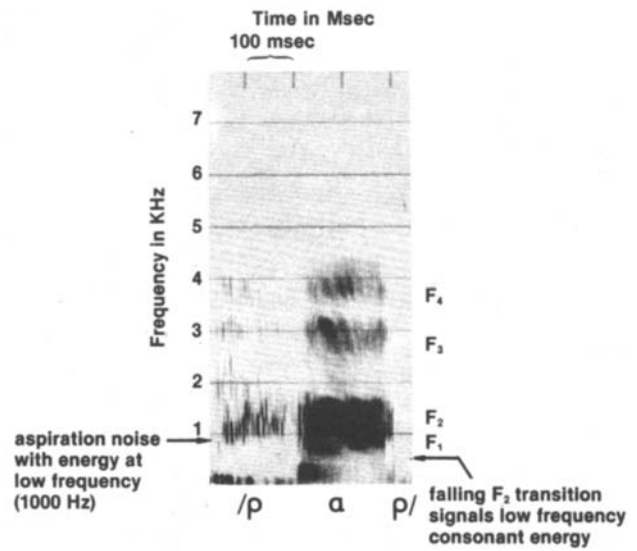
B. Spectrogram

Figure 2–2. Dynamic images of /p/ in word-initial and word-final contexts: "pop." *continues*

Most of the energy for the phoneme /p/ in the initial position is concentrated around 1000 Hz, which is relatively low for a consonant. The first and second formants for the vowel /ɑ/ are somewhat indistinguishable, whereas there is an appreciable distance between the second and third formants. The final consonant /p/ is denoted by the transitional elements of the first two formants of the vowel /ɑ/. Both the F1 and F2 show a downward slant accommodating the final consonant /p/, which has energy in the low-
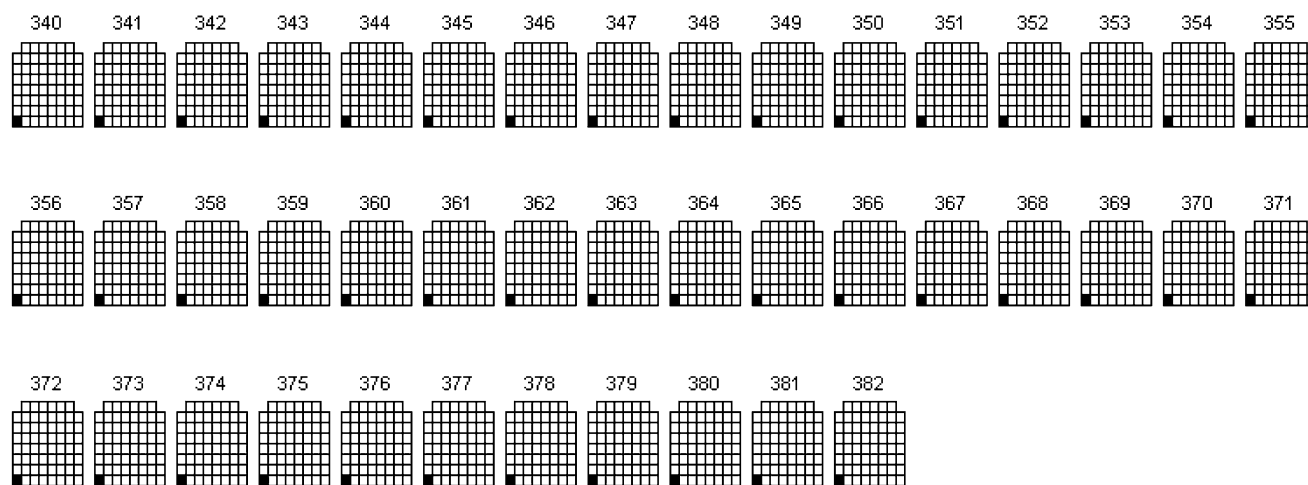
C. Electropalatograph

Figure 2–2. *continued*

frequency areas. Although the articulatory gesture of lip closure is denoted by these formant transitions, there is an absence of any noticeable consonant energy at the final position. This is permissible in English, because stops in the final position in a word may or may not be released. It is clear from this sound spectrogram that, once the information regarding place of articulation (labiality) is conveyed by the downward slope of these formants, it is not necessary for the acoustic information signifying a plosive burst to appear. Although there is a significant amount of aspiration noise present for the initial voiceless-labial-plosive consonant, it is absent in the final position.

Electropalatograph: /pɒp/ "pop"

The dynamic series of EPG palates for the production of the word /pɒp/[1] "pop" is presented in Figure 2–2C. The word "pop" was extracted from the sentence "I see a pop again." Within this sentence context, the word "pop" took 0.209 seconds to produce. Each EPG palate in the dynamic series shows minimal tongue contact palate throughout the whole word, with only one electrode being contacted at the bottom left of the palate. Tongue contact with the palate is not a feature of the production of the consonant /p/, nor the vowel /ɒ/.

/p/ in Within-Word Context: "puppy"

Filmstrip: /ˈpʌpɪ/ "puppy"

The criterion phoneme /p/ is at the initial and medial positions of the word /ˈpʌpɪ/. The first three frames of the filmstrip in Figure 2–3A show the complete lip closure for the initial /p/, and the middle four frames (11 through 14) show the complete lip closure for the medial /p/. Because the vowel /ʌ/ following the initial /p/ is more open than the vowel /ɪ/ following the medial /p/, the lip opening following the initial /p/ is relatively greater and more round (see frame 4) than the flat and small lip opening following the medial /p/ (see frame 15). At both the initial and medial positions, the acoustic tracing of /p/ starts after the lips have opened into the vowel.

Sound Spectrogram: /ˈpʌpɪ/ "puppy"

The phoneme /p/ is presented in both the initial and medial positions in the spectrogram of the word "puppy" (Figure 2–3B). The plosive nature of this consonant is marked by the presence of a plosive burst of energy in both the initial and medial positions. At the initial position, the energy starts at the very base of the sound spectrogram and terminates at over 4000 Hz.

---

[1]/pɒp/ is the Australian English pronunciation of "pop."

A. Filmstrip

B. Spectrogram

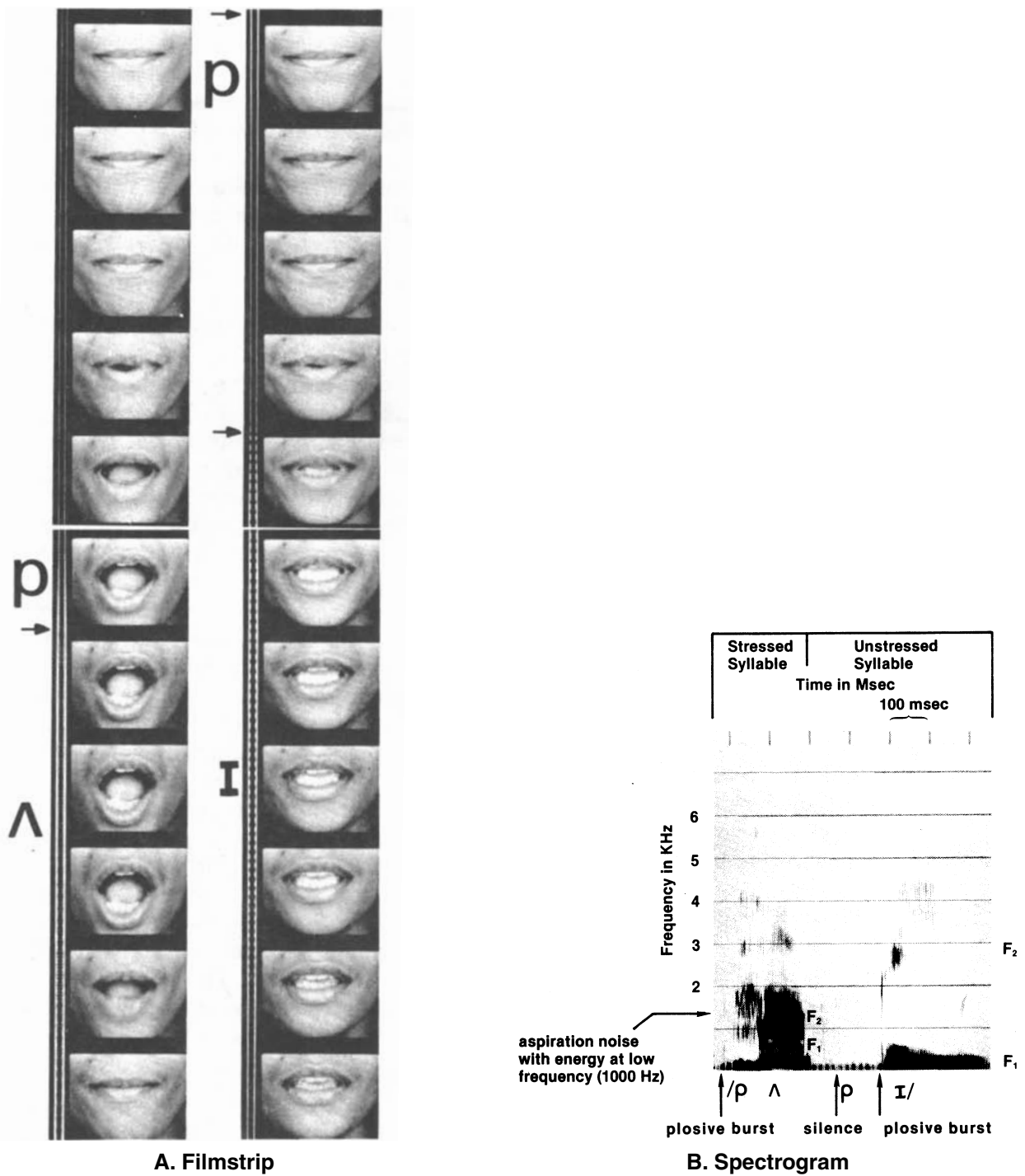**Figure 2–3.** Dynamic images of /p/ in the within-word context: "puppy." *continues*

At the medial position, it again starts at the base and terminates at over 4000 Hz. The difference between the initial position and medial position /p/ is most apparent in the presence of an aspiration noise, which follows the initial plosive burst for approximately 80 msec. This aspiration noise is absent for the medial

Chapter 13

/θ/

"th" (voiceless)

The consonant /θ/ is a voiceless linguodental fricative.

Place of articulation: linguodental

Advancement: front

Voicing: voiceless

Labiality: nonlabial

Sonorancy: nonsonorant (obstruent)

Continuancy: continuant

Sibilancy: nonsibilant

Nasality: nonnasal (oral)

STATIC IMAGES OF THE ARTICULATORY CHARACTERISTICS OF /θ/

The static representations of /θ/ are presented via a photograph, schematic diagram, ultrasound, and electropalatograph images (Figure 13–1).

Photograph
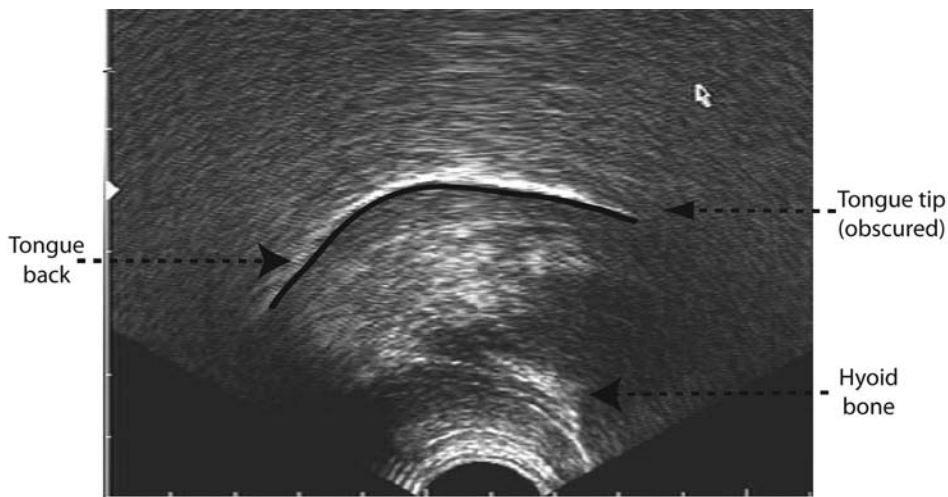
In this photograph (Figure 13–1A), the tongue tip is placed between the lower and upper teeth to channel the airstream to create the low energy sound /θ/.
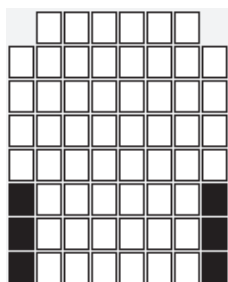
A. Photograph

B. Schematic diagram

C. Ultrasound

Tongue back

Tongue tip (obscured)

Hyoid bone

D. Electropalatograph (EPG) (single frame)

Figure 13–1. Static images of the articulatory characteristics of /θ/.

Schematic Diagram

The lateral view of the production of /θ/ (Figure 13–1B) shows the tip of the tongue placed between the upper and lower incisors. The phoneme /θ/ is a voiceless consonant. Devoicing is symbolized by a (−) in the vicinity of the vocal folds, indicating that in normal speech voicing is attributable to the vocal folds.

Ultrasound

The bright white line on the ultrasound image in Figure 13–1C shows the tongue surface during production of /θ/. The tongue is flat during production of this sound. The tongue tip is on the right and the end is slightly obscured from view because of the acoustic shadow of the jaw (Stone, 2005). Diagonal muscle fibers can be seen below the surface of the tongue.

Electropalatograph (EPG)

The EPG image of /θ/ in Figure 13–1D demonstrates limited tongue/contact during the production of this sound. Tongue contact is located at the posterior lateral margins of the palate. The extent of tongue/palate contact for /θ/ is influenced by the surrounding vowels.

DYNAMIC IMAGES OF THE ARTICULATORY AND ACOUSTIC CHARACTERISTICS OF /θ/

In order to obtain a comprehensive view of the production of this consonant, the dynamic aspects of the production of /θ/ are shown in a filmstrip, spectrogram, and EPG images (Figure 13–2).

/θ/ in Word-Initial Context: "thin"

Filmstrip

The criterion phoneme /θ/ is at the initial position of the word /θɪn/ (Figure 13–2A). In the first frame, the tongue tip can be seen securely placed between the lower and upper teeth. In frame 4 the grip of the teeth begins to relax. Frame 5 shows a complete release of the tongue tip, and frame 6 shows the tongue approximation for the vowel /ɪ/. Not unlike the phoneme /f/, the phoneme /θ/ shows very little energy. The energy is so low that it is undetectable in the sound track presented along the filmstrip. Halfway through the sixth frame the vowel energy begins, the tongue having been released from the interdental (between the upper and lower incisors) position.
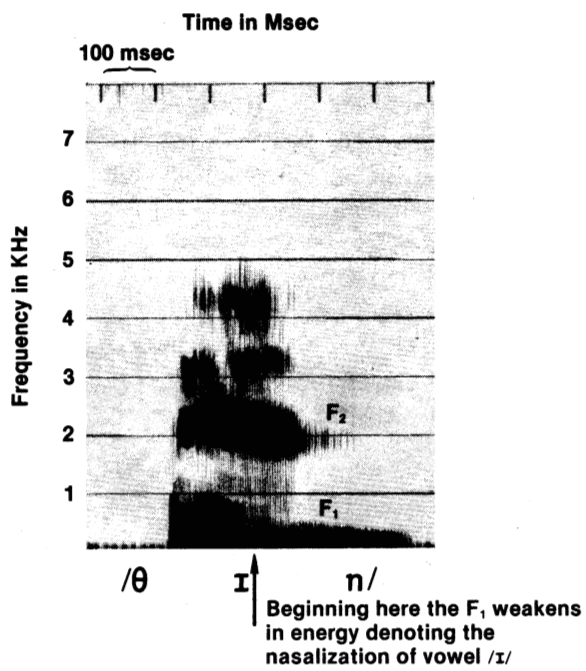
Sound Spectrogram: /θɪn/ "thin"

The consonant phoneme /θ/ is presented at the word-initial position in the context of the vowel /ɪ/ in the word "thin" (Figure 13–2B). The phoneme /θ/, not unlike /f/, shows only traces of energy at a very high-frequency region (around 7000 Hz). The rising F2 transition of the vowel /ɪ/ at 2000 Hz may imply the presence of some acoustic energy at and around that frequency region. Because this phoneme does not manifest itself clearly in the acoustic domain (except when transitions are considered), it is not surprising that problems exist in the processing of this phoneme by children and adults with speech, hearing, and language deficiencies. The influence of the final nasal consonant can be seen midway through the preceding vowel /ɪ/. The weakening of the F1 for that vowel is caused by its nasalization as it is followed by /n/ in the word "thin."

Electropalatograph: /θɪn/ "thin"

Figure 13–2C presents tongue/palate contact for the word "thin." The 63 EPG frames took 0.315 seconds to produce. The initial /θ/ extends from frames 418 to 452, the vowel /ɪ/ extends from 453 to 460 and the final /n/ extends from 469 to 480. The simultaneous sound spectrogram and waveform (not shown in Figure 13–2C) were used to identify the boundaries between /θ/ and /ɪ/ as there was no detectable difference at this point on the EPG palate trace. The EPG frames 461 to 468 represent a period of coarticulatory transition between the vowel and the /n/. The word-final /n/ predominantly has contact with the tongue along the alveolar ridge and the lateral margins of the palate.
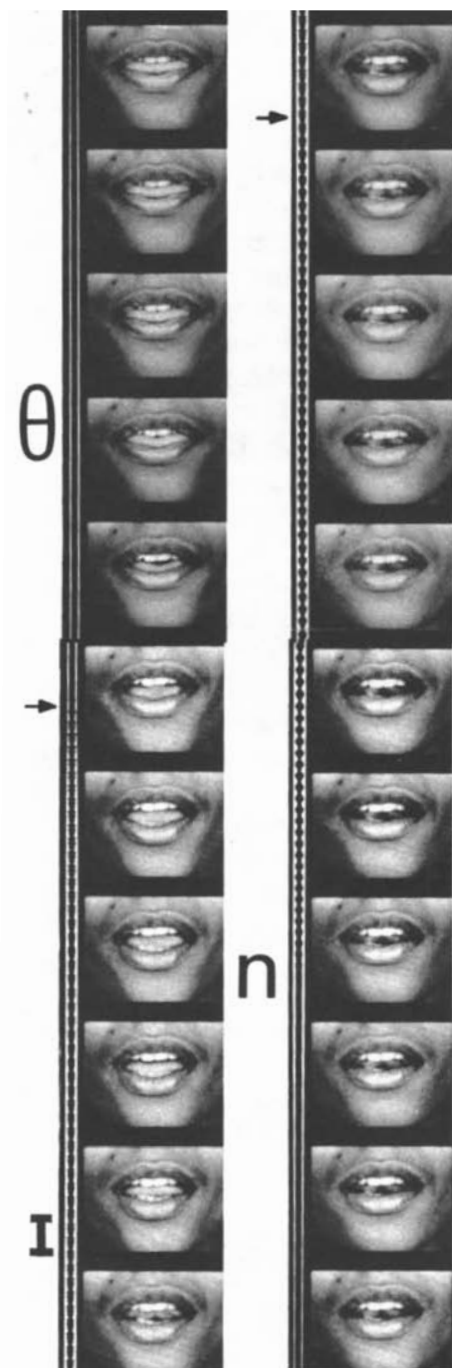
A. Filmstrip **B. Spectrogram**

Figure 13–2. Dynamic images of /θ/ in the word-initial context: "thin." *continues*

Chapter 38

Creating Images of Speech Production

CREATION OF STATIC AND DYNAMIC IMAGES OF SPEECH PRODUCTION

The images displayed in the *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech* represent articulatory and acoustic aspects of consonant and vowel production for English. The majority of images have been specifically created for this book.

Creating Photographs and Filmstrips

The picture filmstrips were made from negatives of 35-mm high-speed motion picture film accompanied by a sound tract. It must be noted that these filmstrips are in absolute continuation. The only reason that some are presented in two, or sometimes three, different strips is the limitation of space in this printed book. One uninterrupted strip would have been too long for any reasonable symmetric presentation. The bottom of the first strip continues at the top of the second strip.

Second, it must be noted that these strips have been edited for presentation. The editing was done only at the beginning and at the end of a word. Because of the long duration of prephonatory articulatory gestures, the part of the filmstrip between the neutral stage of the articulators and the first trace of acoustic energy is not included. However, in each filmstrip, certain amounts of prephonatory and postphonatory articulatory gestures are included to show

that the articulators are active before and after a sound is emitted. In the left-hand margin, phonemes are indicated using the International Phonetic Alphabet and arrows point to the general area of transition between phonemes.

The individual photographs are a single frame selected from the continuous filmstrip to approximately represent the production of a discrete phoneme.

Creating Schematic Line Drawings

Although the high-speed motion picture presentations provide important information regarding the articulatory gestures involved in the production of speech sounds, these pictures primarily emphasize the outside portion of the mouth. Therefore, a line drawing accompanies each module to emphasize the position of the tongue during the production of a particular consonant or vowel. These schematic diagrams are extremely simplified, showing outlines of the lips, tongue, palate and nasooropharynx. At the position of the larynx a plus (+) sign indicates voicing and a minus (−) sign indicates no voicing. The velum is either closed or open, indicating whether the sound is a nasal or oral sound. The tongue shape for each sound was hand-drawn onto the outline of the nasooropharynx by superimposing ultrasound images of the tongue during connected speech. As described below, the selection of each ultrasound image was verified by simultaneous audio and electropalatographic imaging.

Creating Sound Spectrograms

The sound spectrogram is provided for each consonant, vowel, and diphthong. The words presented in each chapter were spoken by a 25-year-old, multilingual female. The speaker learned English simultaneously with two other Indo-European languages. Her dialect has undergone change over time. She spoke General American English at the time of the recordings. The reader should be aware that spectrograms reflect the dialect and sex differences of speakers. Because the speaker in these modules is female, the overall frequency components are elevated. A comparison of the frequency components of speech sounds presented here with those of male speakers generally found in the literature should be made with this consideration in mind. Xue, Hao, and Mayo (2006) described the effect of race on males' vocal tracts and indicated that volumetric differences in vocal tracts may be responsible for formant frequency differences between speakers from white American, African American, and Chinese descent.

Creating Single Electropalatography (EPG) Images

A female researcher who frequently uses EPG in research and teaching recorded the single EPG images for the *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech*. The speaker spoke Australian English, had no history of hearing difficulties, no speech impairment, and no impairment in oromusculature structure or function.

An individual EPG palate for the speech researcher was made from a dental impression of the upper and lower teeth and palate. In this case, a Reading WIN/EPG palate was used that has 62 electrodes arranged in eight rows (see Chapter 1). The dental impression and Reading WIN/EPG palate worn by the speaker is shown in Figure 38–1. The speaker adapted to the palate by wearing it for 1 hour prior to the recording session. Words and phonemes were embedded within a carrier phrase: "I see a _____ again." At the time of the recording session, the speaker simultaneously recorded the EPG, ultrasound,
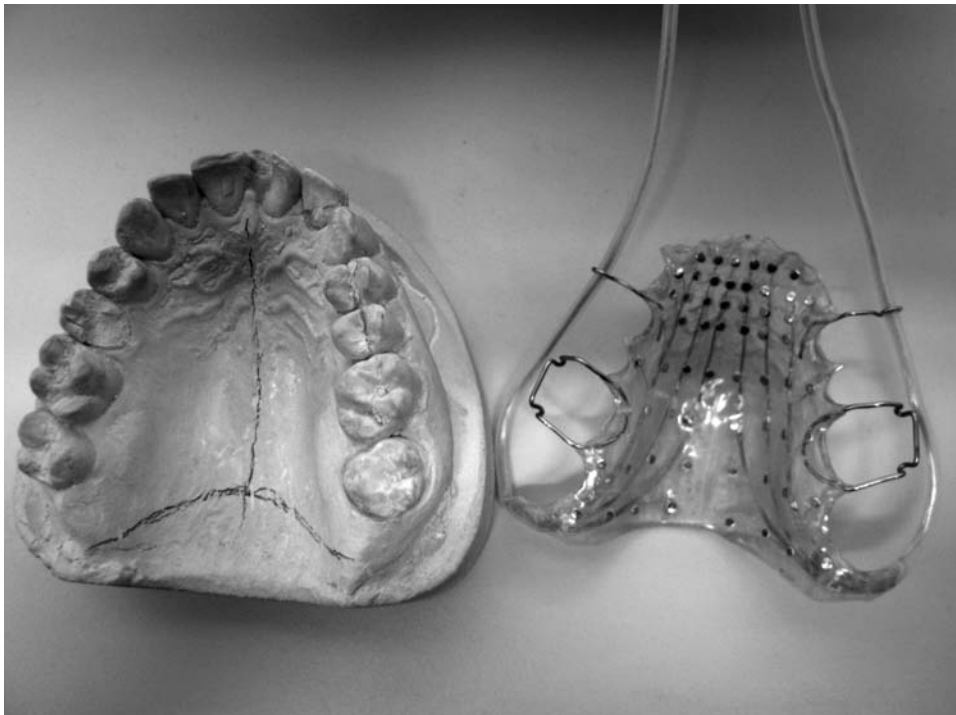

Figure 38–1. Reading EPG palate worn to produce the single EPG images and dental impression used to create the palate.

and acoustic (waveform and spectrographic) data in order to assist with identification of the boundaries and midpoint of the phoneme. Data were collected and analyzed using the Articulate Assistant Advanced (AAA) computer program, version 2.02 (Articulate Instruments, 2006) (Figure 38–2).

Selection of the single EPG frame was based on identifying the midpoint of the phoneme and comparing its similarity with the research data on typical EPG images of maximum contact (e.g., Hardcastle & Edwards, 1992; McLeod & Roberts, 2005; Stone & Lundberg, 1996). In all cases, the selected production was similar (but not necessarily identical) to the typical EPG images published by these researchers. Unlike stylized images of tongue/palate contact, the selected single EPG images within *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech* are at times asymmetric. This is evidence that the images come from real speech. To examine variability in articulation of a specific speech sound, where possible, the single EPG images should be compared with the cumulative EPG productions.

Creating Dynamic Sequences of Electropalatography (EPG) Images

The same speaker who produced the single EPG images also produced the dynamic sequences of EPG images. In each case, the speaker embedded the target words within the phrase "I see a _____ again." Figure 38–3 shows the Articulate Assistant Advanced computer screen for the production of "I see a tat again." The simultaneous display of waveform, spectrogram, electropalatograph, and ultrasound images is apparent. Figure 38–4 shows the complete EPG series of frames for the sentence "I see a tat again" taken from the same recording as shown in Figure 38–3. Due to space considerations, the dynamic EPG sequence of frames shown in Chapters 2 to 27 only include the targeted word selected from the entire sentence. Consequently, Figure 38–5 shows the selected EPG frames to illustrate the word "tat" as presented in Chapter 4. The word boundaries were selected as the frame immediately prior to closure for the initial sounds and immediately following closure for the final sound.

Figure 38–2. Data analysis hardware and software for EPG and ultrasound recordings.

Figure 38–3. Screen image for Articulate Assistant Advanced (AAA) computer program showing simultaneous waveform, spectrogram, EPG, and ultrasound data capture of "I see a tat again."
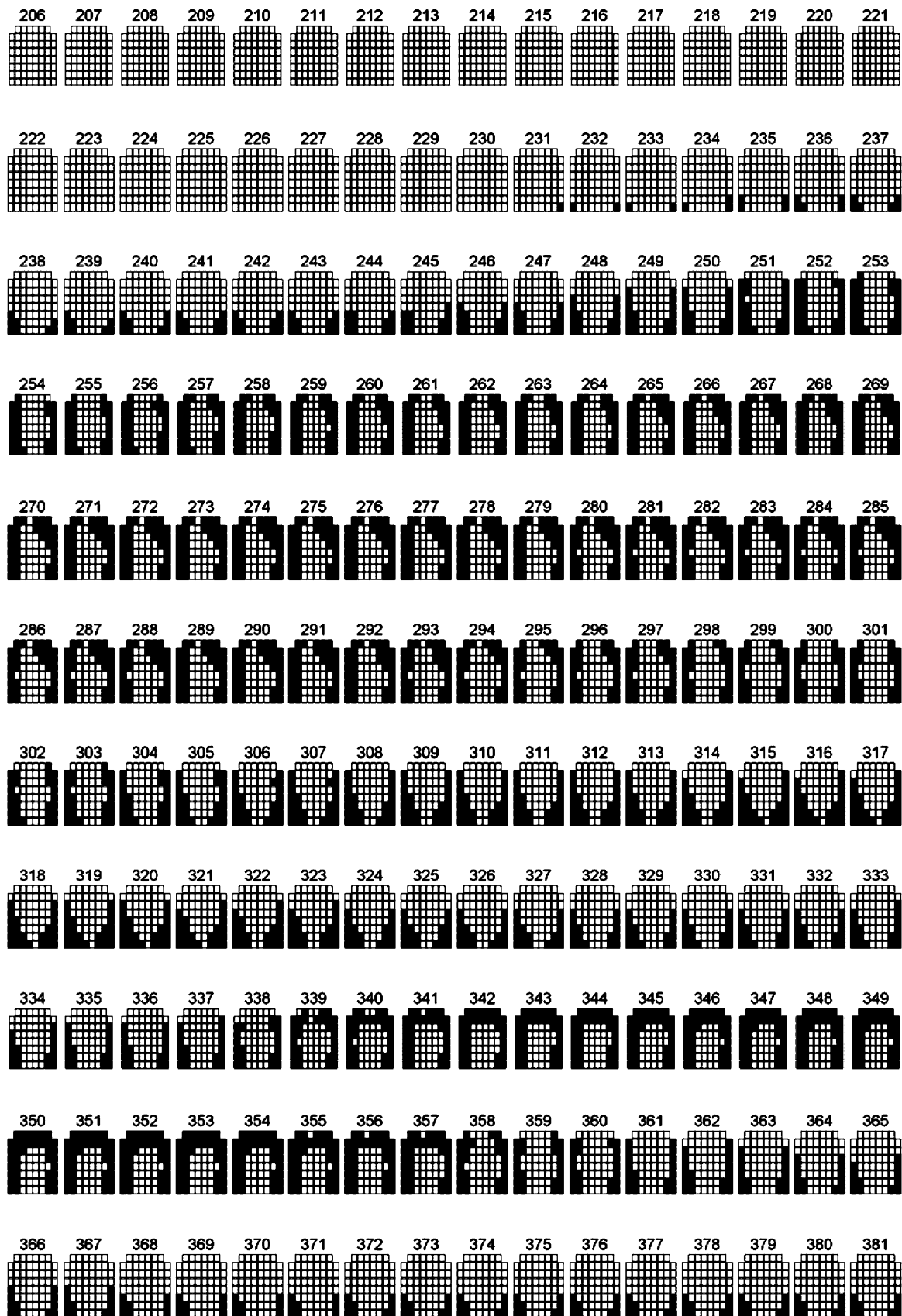
Figure 38–4. Compete series of EPG frames for the sentence "I see a tat again." *continues*
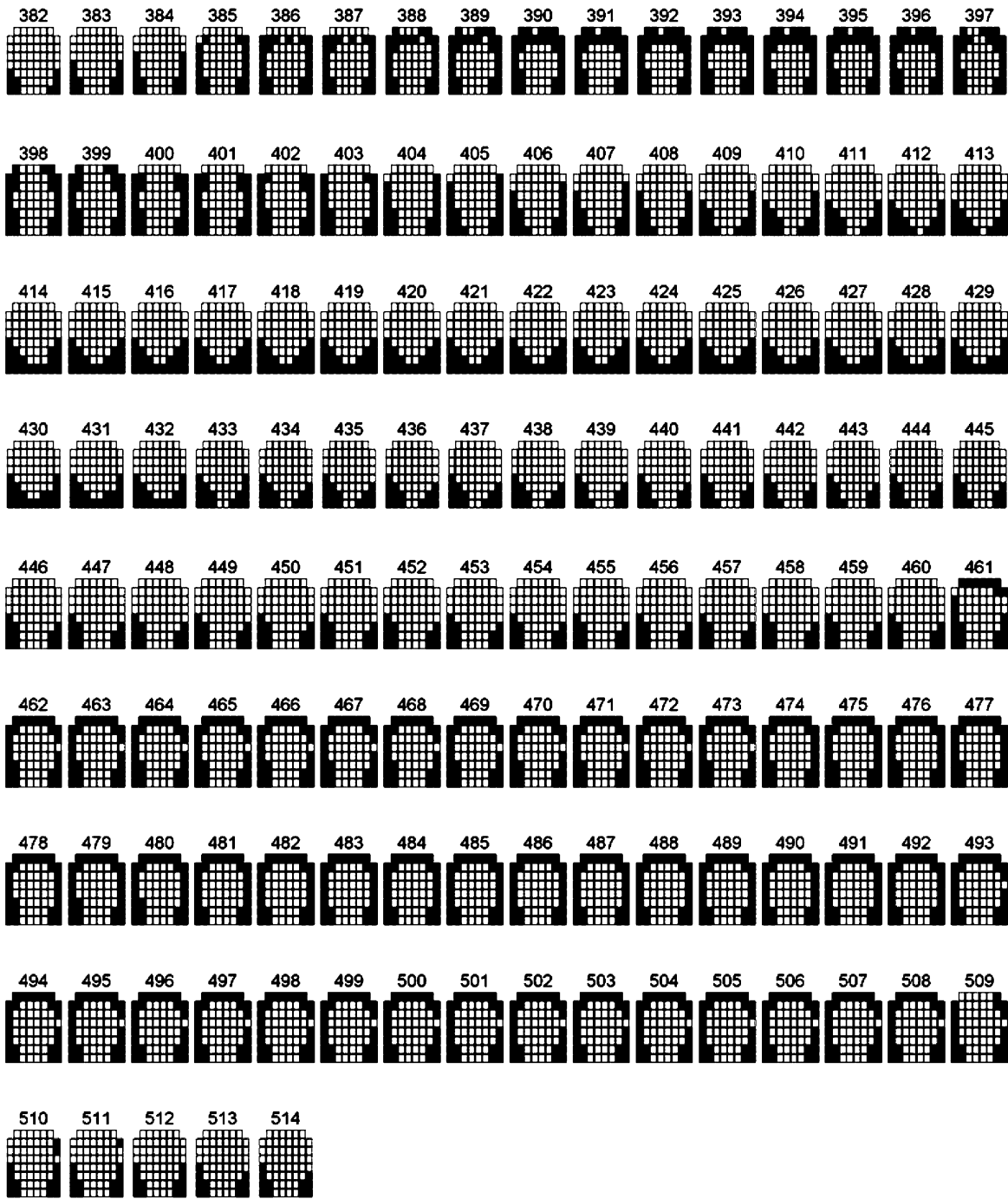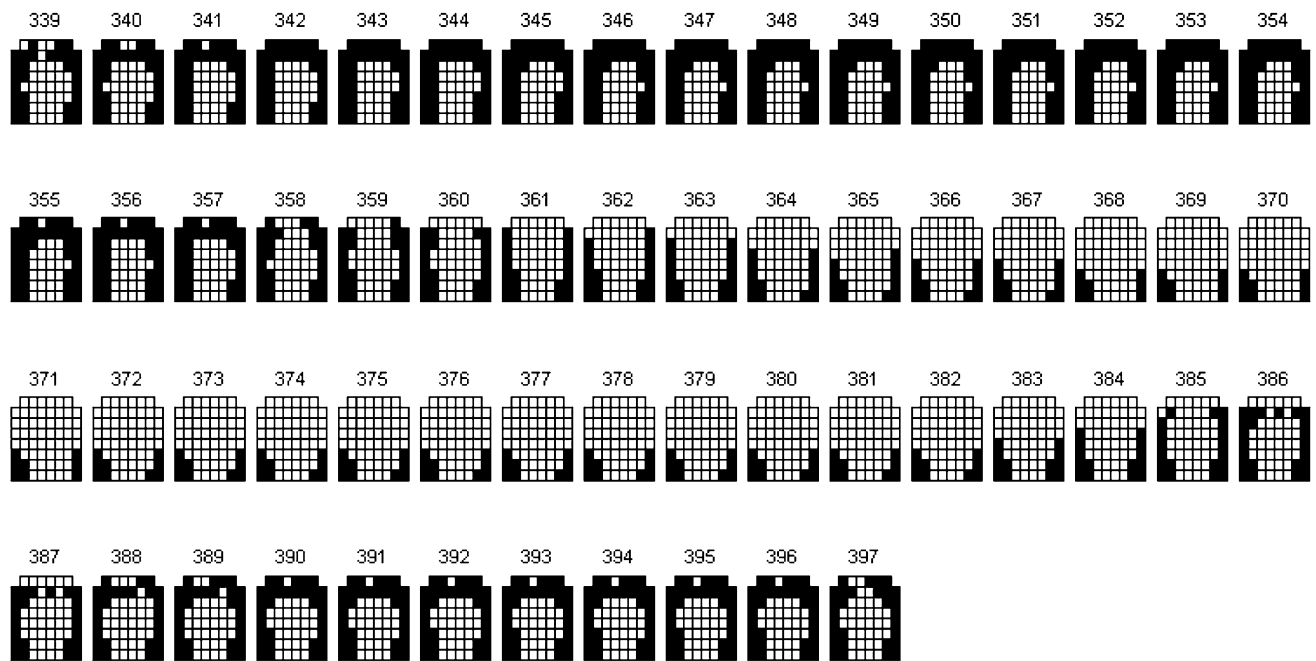
Figure 38–4. *continued*

Figure 38–5. Selected EPG frames for the word "tat."

Creating Cumulative Electropalatography (EPG) Images

The cumulative EPG frames were created as part of a large research project (McLeod, 2003); that has been documented in a variety of research papers and presentations (McLeod, 2006; McLeod & Gibbon, 2007; 2008; McLeod & Roberts, 2005; McLeod & Searl, 2006). The participants were four females (F1–F4) and four males (M1–M4) who spoke General Australian English. None had exposure to speech-language pathology or phonetics. None of the participants reported any significant medical history. Prior to data collection, each participant was assessed and found to have normal results on the following measures: hearing, oromusculature, and ability to produce speech tasks that stressed the phonological system (multisyllabic words, tongue twisters and the Grandfather Passage). The participants produced 13 lingual consonants in a variety of syllable shapes (CV, VC, CVC). The consonants (C) studied were: /t, d, k, g, s, z, ʃ, ʒ, n, ŋ, l, r, j/. Each consonant was paired with a phoneme that occurred at the same alveolar place of articulation, then a phoneme at the opposite place of articulation. The syllables contained the following vowels (V) /i, æ, ʌ, ɜ,

u/ representing the extremes of the Australian vowel quadrilateral. The participants produced the stimuli in the same order and the syllables containing the five vowels were recorded within one phrase, for example, "a neat, a gnat, a nut, a nert, a noot."

To acclimatize to wearing the EPG palate the participants wore a pseudo-EPG palate (the same shape, but without electrodes or wires). On day 1, they wore the pseudopalate for 3 hours over a period of 5 hours and on day 2, participants wore the pseudopalate for 0.5 hours, then the EPG palate (see McLeod & Searl, 2006 for full details). Data recording using the EPG palate commenced immediately following the acclimatization phase on day 2. For the next 4 hours, the participants read a word list of approximately 1000 target words three times over. Productions 2 and 3 of each word were analyzed. These were rated "normal/ typical" by two experienced speech pathologists. If a production was not considered to be "normal" then production 1 was analyzed (this was rare). The maximum point of contact for each consonant was selected as the reference frame (cf. Hardcastle & Gibbon, 1997).

The EPG files were uploaded using ArticAssist v.1.6 (Wrench, 2002). The waveform and audio recordings were used to identify the point where the consonants

were initiated and where they ceased. The frames within these parameters were annotated to represent the series of tongue-palatal contacts for each consonant. The middle 5 frames were then identified and the frame with the greatest amount of contact (maximum point of contact) selected from within these 5 frames.

To exhibit tongue/palate contact patterns for each participant 4 different types of images were created for display in the *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech*:

1. A cumulative display was created by cumulating *all* eight participants' maximum contact frames for their production of all words containing each the consonant.
2. The maximum contact frame from the second production of each word was placed into an extensive table.
3. A cumulative display was created by cumulating all maximum contact frames for *each* participants' production of all words containing each the consonant.

Each of these images was included the *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech* to graphically demonstrate the wide degree of intra- and interparticipant variability for productions of each speech sound.

Creating EPG Images from Others' Research

Throughout the literature, EPG images have been represented in numerous ways. Within each paper, unique varieties of colour (dark to light shading) and shapes (squares, rectangles, and circles) have been used to represent each EPG electrode. To simplify the complexity and individuality of these images and to reduce the need to explain each image separately, all images from others' research have been redrawn using a simple box EPG template. A criterion of at least 67% contact was set for shading of contacted electrodes as black. In almost every instance, it was possible to keep to this criterion and any departure from this criterion is mentioned in the text of the chapter. For example, Guzik and Harrington (2007) reported at least 50% contact in their paper, so it was impossible to use the

67% criterion when reporting their data on three Polish speakers' productions of /s/ in Chapter 15.

There is a substantial body of published literature on electropalatography and its usefulness in understanding typical speech production as well as in the assessment and intervention for people with speech impairment. Due to the size of the corpus of published EPG data, a large amount has not been included in the *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech*. One criterion for the exclusion of data was if researchers had used non-Reading EPG palate designs (such as the Logometrix and Kay designs). These palate designs use a different array and number of electrodes, so it is not appropriate to redraw these images on the 62 electrode Reading EPG palate design. The authors of *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech* would encourage interested readers to read the research from others who have used different EPG palates. Some recommended papers include:

◆ Stone and Lundberg (1996)—typical adults' production of 11 American English vowels + the consonants /l, s, ʃ, θ, n, ŋ/
◆ Stone, Faber, Raphael, and Shawker (1992) – typical adults' production of /s, ʃ, l/
◆ Dagenais (1995) and Dagenais, Critz-Crosby, and Adams (1994)—intervention for children with speech impairment, including lateral lisps

New electropalatography instrumentation is currently being developed. Researchers such as Searl (2003) and Murdoch, Goozee, Veidt, Scott, and Meyers (2004) are developing pressure sensitive EPG palates. Wrench (2007) is developing a new EPG palate enhancing features of the Reading EPG palate used throughout this book. Interested readers can keep in touch with these developments and the new insights that they provide through reading academic journals such as *Clinical Linguistics and Phonetics*.

Creating Ultrasound Images

A female speech researcher who spoke Australian English created the midsagittal ultrasound images of the tongue during speech. In order to stabilize the ultra-

sound probe so that images could be compared with one another a purpose-built helmet was worn (see Figures 38–3 and 38–4) (for further details, see McLeod & Wrench, 2008). Stone (2005, p. 469) indicates that transducer stabilization is important to ensure "intimate contact with the chin and accurate beam direction." The ultrasound images were created simultaneously with the single and dynamic EPG images described above. In Figure 38–6, the ultrasound probe can be seen under the chin and the cords extending from the EPG palate can be seen leaving the mouth and inserting into the multiplexer (box).

In order to select the ultrasound images for inclusion in the *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech*, the audio file, spectrogram, electropalatograph, and ultrasound image were played simultaneously. Each ultrasound image corresponds exactly to the single EPG image to ensure accurate and compatible identification of tongue shape for each consonant or vowel. This simultaneous identification at the point of maximum tongue/palate contact

using EPG and ultrasound is a unique feature of this text and may explain any differences between ultrasound images found in other texts that have not paired ultrasound and EPG imaging.

Rationale for Selection of Speech Models

As far as possible, the authors purposefully used women as the speech models for *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech*. The majority of texts and research papers on articulatory and acoustic properties of consonants and vowels have used examples from males' speech. Kent and Read (2002, p. 189) indicate the need for exemplars from women and children: "The problem is that the research effort given to the speech of women and children has been on a smaller scale than that given to the speech of men. Consequently, there is a continuing need to gather acoustic data for diverse populations." Kent and Read (2002, Chapter 6) provide an extensive discussion

Figure 38–6. Helmet worn to stabilize the ultrasound probe.

of the differences between the acoustic properties of men's, women's, and children's speech. For example, formant frequencies are different for speakers of different gender and ages due to differences in the length of the vocal tract, with an overall trend toward decreasing formant frequencies with age. Females and children tend to have higher formant frequencies than males.

Stone (2005) indicated that the quality of ultrasound images varies according to the speaker. She suggested that images of the tongues of women and children were often clearer than images of men's tongues. She hypothesized that this was possibly due to having smaller and smoother tongue surfaces. She also suggested that people who are thinner have better images than those who are fatter, possibly because fatty tissue in the tongue can refract the sound waves. Furthermore, she suggested that those who are younger have better tongue images than those who are older, possibly because their mouths are moister and they have less fatty tissue.

In comparison, differences between the articulation of sounds are unlikely to be affected by gender. For example, the cumulative electropalatography (EPG) data presented in *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech* have been created by four males and four females. Consideration of these articulatory images indicates that although there are many individual differences, these do not appear to be an artifact of gender. As with texts that primarily use male examples of speech production, the articulatory and acoustic images provided in the *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech* can be used as exemplars for understanding speech production for all people.