UNIVERSITAT DE BARCELONA
Estudios de Fonética Experimental
Journal of Experimental Phonetics
LF Laboratori de Fonètica

Review of "Munro, M. J. (2021). *Applying phonetics: Speech science in everyday life*. Wiley Blackwell".

Lucrecia Rallo Fabra[a]

[a] Universitat de Les Illes Balears (Spain), lucrecia.rallo@uib.es

1. Introduction

I have taught applied phonetics for over twenty years and never have adopted a reference textbook. The options available were either too theory-based (Yavas, 2005) or too focused in the clinical applications (Ryalls & Behrens, 2000). Other pronunciation teaching manuals are based on British English as the target accent (Collins et al., 2019; Gómez González & Sánchez Roura, 2016; Mott, 2005), which would have confused my students since my undergraduate courses are based on American English pronunciation. For these reasons, the moment I had *APSSEL* in my hands, I knew that the manual I had been waiting for so long, finally came out.

The book is clearly written, key concepts are explained in detail making them accessible to the average student and non-specialised audience. But, most importantly, it makes speech science attractive to the reader, something that is not easy to achieve with humanities undergraduates, who usually panic at the first glance of numbers and speech acoustics.

The manual is organized in three sections, the first one is devoted to general concepts in speech science with special focus in speech acoustics. Section II reviews the ontogeny of speech and developmental speech learning in typical and atypical populations. Finally, section III develops current applications of phonetics in various fields of study, among them, speech synthesis and automatic speech recognition, the teaching of second language pronunciation and other recent applications in the film and arts industry. To my knowledge these issues have not been covered in other introductory textbooks published so far.

Each chapter ends with a section that invites the reader to expand his/her knowledge of specific areas with recommended readings. Munro also provides a wide range of engaging activities to promote curiosity, discussion and further research among students.

2. Part I

In the first chapter Munro reviews what makes human language and speech unique compared to other forms of non-communication. He highlights the features proposed by Hockett & Hockett, namely, arbitrariness, discreteness, productivity and duality of patterning. This serves as grounds to distinguish the terms "communication", "language" and "speech". The author then focuses on how technology has changed the study of speech. One example is the widespread use of software such as *Praat* to visualize and analyze the acoustic properties of speech. The chapter ends with a statement of how speech precedes written language. Finally, he shows how the three major branches of phonetics, articulatory, auditory and acoustic, have been extended thanks to artificial intelligence. Instrumental phonetics has greatly contributed to the understanding of many practical applications such as language teaching and learning or voice identification for forensic purposes.

Chapter 2 provides a comprehensive account of the basics of the anatomy and physiology of the speech production mechanisms. Only the essential parts are described, making this section student-friendly.

Undergraduate students may feel overwhelmed with too detailed anatomy descriptions. Speech initiation, phonation and articulation are explained using non-technical words thus making it accessible to the average student.

In chapter 3, Munro justifies the need of a phonetic alphabet for the transcription of speech sounds. The different alphabets used in the world's languages, Roman, Cyrillic or Kanji call for some written convention that represents the sounds of any language regardless of the writing system. The enormous gap between English spelling and sounds further justifies the need to learn phonetic transcription. The IPA alphabet is compared with SAMPA, ARPABET and other transcription systems used in dictionaries such as the Merriam-Webster online dictionary.

Chapter 4 covers the classification of speech sounds from the articulatory perspective with special attention to American English. This is an important contribution of the textbook, since most phonetics textbooks in the market are based on British English. The articulatory properties of the English vowels and consonants are condensed but no details are missing. The last section includes examples of speech sounds that do not occur in English but are part of the sound inventories of other languages such as the Spanish trill or the Arabic pharyngeal consonants.

In chapter 5 the working definitions of lexical stress, rhythm and intonation are provided and illustrated with examples of real utterances and *Praat* screenshots. Understanding the cross-linguistic differences and acoustic correlates of stress are often challenging for the average student. The *Praat* screenshots comparing the differences between stressed and unstressed syllables are informative.

3. Part II

The first chapter in this section tackles the controversial issue of whether speech is exclusive of the human species or if other species exhibit human-like language capabilities. In an evolutionary line, Munro provides evidence that human speech tracks back to the homo sapiens. Although some primates possess vocal anatomies similar to the human vocal tract, it is uncertain whether the first hominis' cognitive abilities were comparable to those of present-day humans. Unlike skeletons, soft structures do not fossilize. A review of several studies experimenting with animal species show similarities with the human language. Some examples include a parrot's ability to imitate human speech, a primate's vocalization to signal threat or a dolphin's response to recordings of its own signature whistles.

In chapter 7 Munro goes through the development of speech from birth to adulthood. From the anatomical perspective, we learn the commonalities between babies and chimpanzees as far as the location of the vocal tract is concerned. Interestingly, both are born with larynxes but these lower for children but not for chimpanzees. Age-related as well as sex-related differences are also reported. As for perception, we learn that prenatal babies are sensitive to many speech features. Measuring the heart-rate, researchers know that they can discriminate the mother's voice from other voices or react to changes in pitch. As early as four months, infants can categorize voicing contrasts such as /p/-/b/. When infants approach the first fear of life, their perceptual systems become attuned to the language of the environment, ignoring phonetic differences that are not relevant to the L1. This language-specific attunement is also found in speech production. Acoustic analyses of the cry of newborns show cross-linguistic differences in pitch alignment and intonation.

L1 experience further shapes infants' preference for place of articulation. The empirical evidence shows that the first articulations are produced at the larynx regardless of the L1. As they approach the first year of life, children from Arabic households continue to produce pharyngeal consonants because these are the sounds of the environment. In contrast, English-speaking toddlers show a preference for coronal articulations. There is no fixed pattern of order of acquisition of speech sounds. On the contrary, individual variability is often the norm. In some cases, consonants such as /r/, /l/ or /s/ might not emerge until the child reaches school age at five or six years.

Adult speaker identity is the result of organic and learning influences. It follows that pitch and voice quality are linked to the size and texture of the vocal tract, its shape, length and the nature of the tissues inside it. In turn, social experience shapes the differences between speakers of the same language in terms of the different accents. Speakers of any language can switch accents if they change residence and are exposed to different varieties of the L1. Aging affects voice quality due to atrophy of the vocal structure. These changes are more noticeable in women than in men due to the hormonal changes that come along with menopause. Other vocal changes among adults may be caused by overuse as it is the case of professional singers, who might develop vocal disorders as they mature.

In chapter 8, the author approaches the most common speech disorders and their etiology. Cleft lip and/or palate cause hyper nasalized speech but fortunately it can be remediated through surgery early in life. Fluency disorders characterized by the involuntary repetition of phones, syllables and words have the highest incidence according to the US health authorities. Developmental fluency usually resolves itself within six months without intervention. Specific intervention methods based on breathing, control of tempo are the most commonly used to treat stuttering among adult populations. Other severe speech disorders result after laryngectomy, the partial or total removal of the larynx caused by cancer or neck injuries. Laryngectomees are forced to use alternative voice sources such as esophageal speech, external vocal prosthesis or laryngeal transplantation.

Aphasias are caused by brain injury and can affect four language skills: speech production and perception, reading and writing. Broca's aphasia is characterized by the inability to produce speech due to damage in the frontal lobe. In contrast, Wernicke's aphasia is linked to a lesion in the temporal lobe and it results in fluent and grammatical speech albeit incomprehensible by the listener.

Other motor speech disorders such as dysarthria and apraxia affect the coordination of speech muscles and can be congenital and acquired. The most common etiologies include cerebral palsy, lateral sclerosis and stroke. Finally vocal abuse can trigger some voice disorders such as dysphonia. This condition is common among singers and actors or as a consequence of smoking and/or aging. The chapter ends with a note about the ASHA (American Speech-Language-Hearing Association) and the role it played after the II World War, treating soldiers who suffered from language impairments after brain injury. Today, speech therapists are responsible for counseling, speech and language assessment, post-surgical therapy to restore a client's impaired language skills.

4. Part III

Chapter 9 opens section III of the book devoted to applied phonetics so to speak. The author reviews some of the early attempts to build "talking heads", considered the precursors of today's speech synthesis or process of creating spoken utterances without a human tract. Reference to these devices can be found in Cervantes' writings but the earliest prototypes date from the 18th and 19th centuries. Von Kempelen needed 20 years to develop his talking machine which could synthesize French and Latin. In the 19th century, the understanding of the tube-like behaviour of the human tract pushed other scientists to create other talking robots, such is the case of Faber's Euphonia, which was displayed in his NY museum.

Synthesizers had their golden age at the beginning of the 20th century thanks to some scientific advances such as the application of x-rays to speech physiology. Advances in speech acoustics also made possible the electronic generation of speech from scratch. American telephone companies showed a growing interest in applying speech synthesis for customer services. This scenario favoured the creation of the *Voder*. Researchers at Haskins Labs used spectrographic patterns to create consonants and thus were able to identify the chief acoustic differences between stop consonants. In the 1980s Dennis Klatt, a professor from the MIT, developed the first text-to-speech synthesizer to help the blind by reading printed material aloud.

Differences between natural and synthesized speech are provided and illustrated with sound waveforms and speech samples available at the *APSSEL* website. The challenges of text-to-speech synthesis are discussed. These typically involve capturing the pitch changes that characterize human speech and somehow need to be implemented in commercially-available systems such as *Siri* or *Alexa*. Further to this, the opaque nature of English spelling poses many difficulties for TTS (text-to-speech) systems, which can be partially solved introducing spelling to sound rules into the system.

In chapter 10 Munro introduces the field of forensic linguistics and how speech science can help solve crimes thanks to earwitnesses, who can identify suspects providing descriptive information about a speaker's voice in terms of accent, pitch or an identifiable speech disorder. Throughout a historical review of real cases, the reliability of earwitnesses is questioned. Expert speaker identification (SID) is the updated method of former voice line-up. Naïve/expert judges are requested to rate the similarity between the voice of the perpetrator (unknown) and the suspect (known) on the basis of various dimensions, namely, pitch level and variability, voice quality and speech errors among many others. In any case, this evidence on its own is not sufficient to convict a suspect in a court case. Additional evidence is required. Unlike fingerprints, a speaker's voice cannot be labeled through such a thing as a "voiceprint". The reason is simple: voice is the result of combining organic factors (length of the larynx and vocal folds) and learned factors shaped by a speaker's interaction with other members of the community. Acoustic measurements can also be used for voice comparison, specifically the mean and standard deviation F0 can help identify a given speaker. Voice breaks are also informative in that they are indicative of a creaky voice. Other speaker idiosyncrasies include the measurement of harshness and nasality.

Computer-based matching of voices or automatic speaker identification have evolved thanks to the advances of artificial intelligence. However, the applicability to SID is limited. These systems use statistical analysis to estimate the probability that two voice samples belong to the same individual. Unfortunately, many average speakers share the same speech features (pitch, resonance). One of the reasons that make SID challenging is the use of mechanisms for disguising the voice, such as changing pitch, denasalizing speech or adopting a foreign accent. These, along with the poor quality of the recordings can make SID really challenging.

Chapter 11 deals with one of the most popular topics in applied linguistics, namely, pronunciation teaching. The first attempt to aid the general public to speak correctly dates back to the 1600s. *The Vocal Organ* was the first attempt to teach the pronunciation of English vowels and consonants. Owen Price's classification of consonants as "throat" and "lip" letters was far from accurate but he managed to sell a considerable number of exemplars. The play *Pygmalion* by B. Shaw and its subsequent adaptation to the film *My Fair Lady* reflected the attitudes of 20th century British society towards non-standard accents and their stigmatization. These attitudes have also been found in the US towards Southern American English. In the 20th century, the early interest of the phoneticians David Abercrombie and Pierre Delattre in applying the principles of phonetics to teach languages, set the groundwork to develop the IPA and phonetic transcription.

Accented L2 speech is the norm for those who have learned an L2 past early childhood. According to Munro, foreign accent should not be regarded as negative or as failure to learn as long as comprehensibility is not compromised. Interestingly, a native listener is able to detect a foreign accent even in brief speech samples as short as 30 milliseconds. Some researchers have viewed this activity as an evolutionary advantage that allows the members of a community detect possible threats to their safety from outsiders.

The age factor has great impact on foreign accents. Infants' perceptual attunement to the sounds of the L1 after the first year of life has the cost of losing sensitivity to L2 sounds. Some of these capacities can be restored later in life through perceptual training methods, but adults rarely reach the standards achieved by infants. The most common segmental errors produced by non-native English speakers are reviewed. Reference is made to the /r/-/l/ confusion by L1-Japanese learners, substitution of /ɪ/ with /i/ by L1-French or L1-Italian learners. These errors can be attributed to the differences between the L1 and L2 sound inventories, as it was suggested by the *Critical Period Hypothesis.* However, this theory does not tell the whole story, especially when it comes to explain the individual differences between learners.

Errors involving suprasegmental aspects of speech reported in the literature include the addition of epenthetic /e/ to facilitate the production of clusters /st/ and /sp/ by L1-Spanish learners, errors in lexical stress placement by L1-French speakers. English rhythm can also be challenging for French or Japanese speakers. Unlike native speakers, they give the same prominence to all the syllables and thus fail to reduce the stressed ones. The inappropriate use of intonation patterns may cause misunderstandings of the speaker's attitudes. The four dimensions used to measure L2 speech, namely, foreign accent, intelligibility, comprehensibility and fluency are discussed. Munro provides empirical evidence showing that these dimensions are not necessarily interrelated, a given speaker may be perfectly intelligible yet heavily accented.

Pronunciation teaching, along with grammar, vocabulary or writing should be an integral part of L2 learning. Reasonable standards in L2 pronunciation guarantee successful communication. The pros and cons of the "nativeness" and "intelligibility" principles are discussed. Munro advocates for the latter, since nativeness is an unattainable goal for the vast majority of learners. The role of functional load is also discussed, segmental errors involving a high functional load, such as the /r/-/l/ contrast in L1-Japanese English seriously compromise intelligibility because the listener may confuse word pairs such as *write-light* or *berry-belly.*

The final section of this chapter lists the different teaching practices used by teachers to correct pronunciation: recasting, shadowing and computer-assisted pronunciation learning. Technology has allowed the development of more customized techniques such as *High-Variability Phonetic Training (HVPT), ASR* and gamification, allowing more individualized teaching practices and making pronunciation learning more enjoyable for young learners.

Chapter 12 deals with two interesting applications of phonetics: accent coaching and singing. Knowledge of phonetics allows pronunciation coaches to modify the accents of actors in the film industry. The terms accent and dialect are explained with special emphasis on regional accents. The pronunciation differences between the two major accents, British and American English are outlined. Two different approaches to accent coaching are described: the "prescriptivist" method led by Edith Skinner and the "intelligibility" principle proposed by Dudley Knight. The former advocated that the Mid-Atlantic accent was the model professional actors should follow. In contrast, Knight advocated for a more intelligibility-based model. The techniques used by accent coaches range from analytical techniques, giving explicit instructions on how to produce particular segments and words, to more holistic techniques. The latter are necessary to teach voice qualities such as breathy, nasalized or foreign-accented speech. Interestingly, IPA transcription is rarely used by accent coaches, instead they resort to orthographic renditions, a more practical approach for their non-linguistically trained clients.

The last section of this chapter focuses on the comparison of speech and singing, specifically pitch variation in each sound modality and the capacity of voice professionals to sing above the frequency of the orchestra. The acoustic characteristics of singing or "singing formant" are clearly illustrated with spectral displays. Finally, a brief review shows the reader how contemporary pop music owes a lot to digital processing techniques that can correct "off-key" pitch errors or edit voices to create distortions or robotic effects.

Chapter 13 starts with an introduction to audio-visual speech perception, highlighting how the weight of visual cues is crucial for effective communication. Visual information from speech gestures has been used in the animation industry since the early 20th century. The term *viseme* was introduced to name the animation minimal facial patterns representing a particular posture for a given vowel or consonant. Disney studios animated films were based on these patterns. More recently, the development of artificial intelligence has given rise to more efficient animation methods. The process works as follows: a typical utterance is first phonetically transcribed automatically; the resulting IPA transcription is subsequently matched to the corresponding visemes for the animated face. Visemes have also been improved to reflect the dynamic characteristics of speech.

The chapter ends with some curiosities such as the ability of ventriloquists to produce speech with virtually no lip movement. This can be achieved by modifying pitch and raising the larynx to enhance resonance. A brief reference is made to *conlangs* or constructed languages, that is, artificial languages created by writers/ film directors for fictional purposes. The Elvish languages from *The Lord of the Rings* or Klingon from *Star Trek* have their source of inspiration in many indigenous languages from the Americas. The phonetic inventories of these languages include articulations that are rare in the world's languages such as retroflex and ejective consonants.

Chapter 14 covers the contributions of phonetics in the business world. The first sections include a brief historical review of the development of ASR systems for commercial use. The first attempts to apply this technology in customer-support phone calls emerged in the 1950s and were sponsored by the Bell Telephone company. *Audrey* could recognize numbers from one to ten spoken by male speakers. The system was based on a pattern-matching procedure that used formant frequencies to recognize individual sounds. Twenty years later, *Harpy* could recognise full sentences produced by multiple speakers thanks to a probabilistic modeling method, which predicts word identity from context. The impact of ASR in everyday life ranges from dictation software that converts speech into orthography to phone apps that function as pronunciation coaches, such as *ELSA Speak.*

The following sections refer to how some speech features can convey certain feelings and attitudes in the consumer society. Some research studies in sound symbolism suggest that the front vowels are associated with small-size objects, whereas back vowels better represent large-size items. Interestingly, fast-rate low-pitched voices are judged to be more attractive than slow-rate high-pitched voices. These features along with accent have been found to influence consumers' preferences for certain commercial brands.

In the closing chapter, Munro raises the reader's awareness about some ethical issues of speech and language experts and their questionable practices. For instance, voiceprints or spectrograms were used as proof to convict a suspect in a US court case. Luckily, the defense provided proof that a speaker identity could not be based on this sole evidence. *Vocal stress analysis* systems advertised through the internet claim to determine whether an individual is telling the truth through some acoustic properties of speech. Surprisingly, this software is widely used by police forces in North America even though researchers have warned the public about the false-positive rate of these systems. Other issues that raise controversy include the fraudulent manipulation of speech to fake the voices of celebrities or politicians and the growth of online services that claim to eradicate their potential clients' foreign accents and achieve nativelike standards.

Acknowledgements

This book review was prepared with the financial support of research grant PID2020-117804GB-100 AEI/FEDER from the Spanish Ministry of Science and Innovation.

References

Collins, B., Mees, I. M., & Carley, P. (2019). *Practical English phonetics and phonology: A resource book for students.* Routledge.

Gómez González, M. A., & Sánchez Roura, T. (2016). *English pronunciation for speakers of Spanish: From theory to practice.* De Gruyter Mouton.

Mott, B. (2005). *English phonetics and phonology for Spanish speakers*. Publicacions de la Universitat de Barcelona.

Ryalls, J., & Behrens, S. (2000). *Introduction to speech science: From basic theories to clinical applications.* Allyn & Bacon.

Yavas, M. (2005). *Applied English Phonology.* Blackwell.