Speech Acoustics of the World's Languages

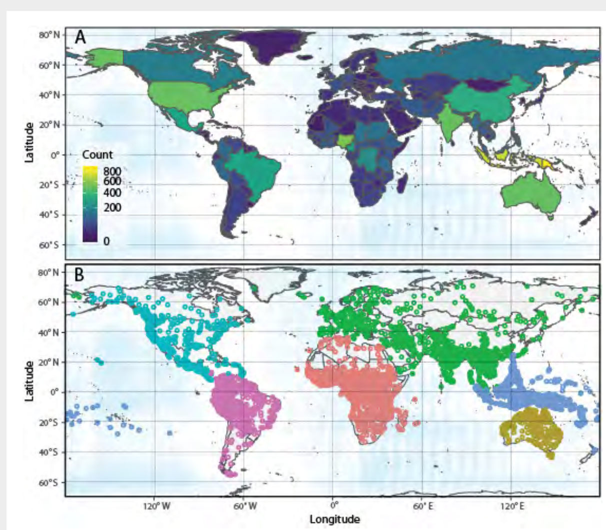
Benjamin V. Tucker and Richard Wright

Depending on how one counts, there are more than 7,000 languages in the world (Eberhard et al., 2019). Many languages contain unique and interesting sounds and combinations of sounds that speakers produce to convey meaning to listeners. A language may share similarities with other languages on some dimensions while also having differences along other dimensions. Because of this linguistic diversity, it is important to the understanding of speech communication, and, more specifically, sound production in the world's spoken languages, to sample the sounds of language as broadly as possible. In the present article, we briefly discuss the diversity of the world's languages and speech sound production mechanisms. We also discuss the importance of documenting the acoustic characteristics of these sounds and the role of linguistic extinction on our ability to adequately sample the sounds of the world's languages.

Over the last century and a half, speech researchers have developed a reasonably good understanding of how speech sounds are produced in the vocal tract. Given this, we might assume that sampling any single language, or even a handful of languages, might be sufficient for understanding speech sounds. However, even a language like !Xóõ (Traill, 1985) that is spoken in Botswana by about 2,000 speakers (Eberhard, et al., 2019), with 58 consonants, 31 vowels, and 4 tones, covers only a fraction of the attested speech sounds in the world's languages. Similarly, as pointed out by Ian Catford (1977) and Björn Lindblom (1990), from an *anthropophonic* (human sound) perspective, the vocal tract is capable of producing a much wider variety of sounds than are used in human language (e.g., beatboxing; Proctor et al., 2013) because linguistic sounds are constrained to be efficient vehicles for communication. Therefore, as scientists, it is important that we sample languages broadly rather than relying on a handful of well-documented or closely related languages from restricted geographic distributions. For historical and demographic reasons, linguistic diversity and the research sampling of languages is not evenly distributed across the globe.

The maps in **Figure 1** illustrate the linguistic diversity of the world and both maps illustrate the uneven global distribution of languages. **Figure 1A** is a heat map of each country in the world showing the number of languages spoken in that country, including both indigenous and immigrant languages (Hammarström et al., 2019).

Figure 1. ***A:*** *world heat map of individual countries.* ***Colors****, total number of languages reportedly spoken in each country as per Glottolog (Hammarström et al., 2019).* ***B:*** *world map of language location.* ***Circles****, latitude and longitude associated with an individual language as reported by Glottolog;* ***color*** *is associated with one of six regions (North America, South America, Eurasia, Africa, Australia, and Oceania). Where circles become difficult to distinguish represents a dense linguistic region.*



This article derives from a special issue of The Journal of the Acoustical Society of America *on the Phonetics of Under-Documented Languages that was edited by Benjamin V. Tucker and Richard Wright. You can see all articles in the issue at acousticstoday.org/pudls.*

https://doi.org/10.1121/AT.2020.16.2.56

Figure 2. *Abbé Rousselot (1846–1924) with a kymograph. Available at bit.ly/2wrHMRV.*

Figure 1B takes the latitude and longitude data associated with each of the languages in the Glottolog dataset (a comprehensive catalogue containing basic descriptive information for many of the world's languages; Hammarström et al., 2019) and plots it on the map.

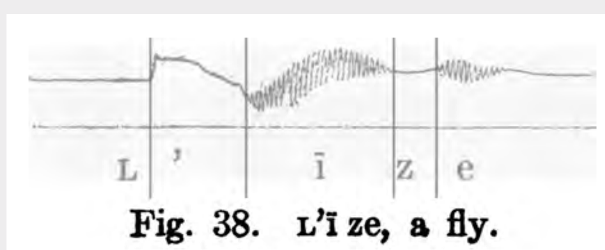
Both maps illustrate the tremendous linguistic diversity across the world. Nearly all inhabitable regions have at least several languages. As illustrated in **Figure 1B**, many regions have a high density where the languages are plotted with overlapping circles, causing those areas to become solid. In many cases, the dense regions in **Figure 1B** contain languages that have been studied the least, leaving ample opportunity for acoustic studies. For example, Equatorial Africa, India, Papua New Guinea, and the northern part of South America are solidly covered in circles, which indicates a very high language density.

A Historical Perspective on Speech Sound Studies

Early phonetic documentation of the world's languages was relatively broad and considered many languages. By way of illustration, in the late 1800s and early 1900s, Abbé Rousselot (1897) adapted the kymograph (**Figure 2**), invented for medical research, to his research on speech. The kymograph was used to record both nasal and oral airflow and was also able to detect vocal-fold vibrations (laryngeal activity). Not long after Rousselot's publications, P. E. Goddard took a kymograph into the field where he recorded two Athabaskan languages: Hupa (Goddard, 1905), spoken in northwestern California, and Dene Sųłiné (Goddard, 1912), spoken in north-central and northwestern Canada and which is reported to currently be spoken by about 10,700 speakers as their native language (Eberhard et al., 2019). As is illustrated in a recording of Dene Sųłiné published in 1912 (**Figure 3**), the kymograph recorded not only oral airflow but also the vocal-fold vibrations or voicing of the speech. In **Figure 3**, the speech sounds are demarcated in [tł'i:ze], "a (horse) fly." In the phones with high-amplitude voicing (vocal-fold vibration), the regular vibrations of the vocal folds can be seen. This early work indicates an interest and desire by early speech researchers to document the sounds of the world's languages and to describe the unique aspects of these sounds, although a bias often remained toward languages that were easily accessible to the researcher.

By the late 1920s, acoustic studies of speech had become more common, and the first decade of *The Journal of the Acoustical Society of America* (*JASA*) saw 17 papers on the acoustics of speech and speech production. Of these, 12 were either about English or used English exclusively as data for a study and the others used nonlanguage vocalizations. The pattern of focusing largely on English continued even as acoustic studies of spoken language became more widespread with the release of the spectrogram in the 1940s (Koenig et al., 1945; Potter, 1945). Thus, as acoustics became more widely used in linguistic and psychological research after World War II, a small

Figure 3. *Goddard's (1912) kymographic tracing of airflow of a speaker producing the word [tł'i:ze], "a (horse) fly" in Dene Sųłiné, with segmentation of individual phones added.* x*-axis, time;* y*-axis, amount of airflow. Vocal-fold vibration is also shown in the movement on the* y*-axis.*



handful of languages (e.g., English, French, German, Dutch, Japanese, and Chinese) came to dominate most publications investigating the acoustic characteristics of speech. A notable exception for *JASA* is a paper in a special volume on communication from 1950 in which the author, John Lotz, points out the need for studying speech acoustics from a cross-linguistic perspective (available at asa.scitation.org/toc/jas/22/6).

"*Every speech event belongs to a definite language. Any speech analysis that disregards this fact...will lack adequate principles for the classification and description of the complexities of speech*" Lotz, 1950, p. 712.

Although most of the papers in this special volume were theoretical in nature and therefore contain very little actual acoustic information and no acoustic investigations about specific languages, Lotz's point is well founded. Despite the call for more diversity in the acoustic studies of languages in 1950 and despite a renewed interest in languages of the world among linguists, the bias toward relying on a handful of languages persists even today.

This general lack of acoustic description and research on underdocumented languages in *JASA* and in other journals inspired us to host two special sessions on the phonetics of underdocumented languages at Acoustical Society of America meetings (in Salt Lake City, UT, May 2016, and Victoria, BC, Canada, November 2018) and to organize a special issue on the phonetics of underdocumented languages (Tucker and Wright, 2020). One goal of the special issue is to increase the number of underdocumented languages described in *JASA*. In the special issue, there are descriptions of aspects of 25 different underdocumented languages from 5 different continents.

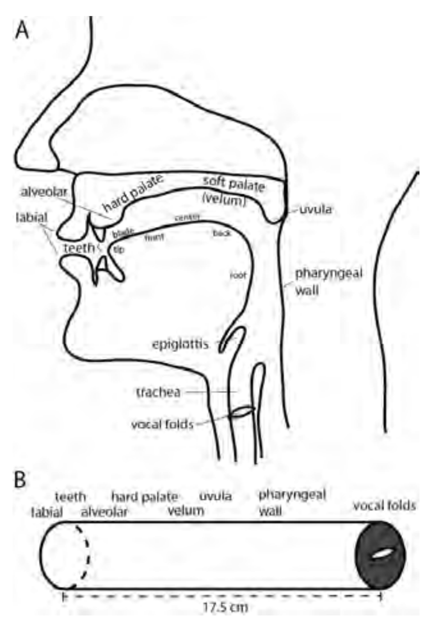
Acoustics of the World's Languages

Many sounds have been described phonologically in linguistic grammars of languages, descriptions that explore and explain the patterns of a given language, although the phonological section in these grammars typically makes up a very small portion of the grammar. Most linguistic grammars use impressionistic methods where essentially the researcher writes down what they think they hear. In these grammars only a fraction of the described sounds have been examined in phonetic, and particularly acoustic, detail. PHOIBLE is a database of speech sounds that lists 3,183 speech sounds in 2,186 languages (Moran and McCloy, 2019). Many, maybe most, of these sounds and sound combinations have not been described acoustically. First, we follow the International Phonetic Alphabet (2018; see at acousticstoday.org/ipa-chart) conventions for transcribing speech sounds and indicate that these are speech sounds using square brackets on either side of the sound. Then, we briefly describe some of the unique acoustic characteristics of speech sounds from several different languages that are acoustically underdocumented.

Source-Filter Model of Speech Production

A simplified way of modeling speech sound production is using a source-filter model (Chiba and Kajiyama 1945; Fant, 1960) with a source (e.g., vocal-fold vibration or aperiodic turbulence) that is filtered by the shape of the vocal tract (**Figure 4A**). A way to realize this model is by using a tube as the filter with a source at one end of the tube (e.g., **Figure 4B**). There are a variety of possible sources at different points along the vocal tract. The easiest vocal tract configuration to start with is a vowel, where vocal-fold vibration creates a complex

Figure 4. ***A:*** *midsagittal view of the human vocal tract with important places of articulation labeled.* ***B:*** *neutral closed-open tube with labels indicating the approximated places of articulation. The closed end is the dark oval on the right at the vocal folds and the open end is located at the labial end of the tube. See text for further explanation.*



sound that is filtered by the resonant characteristics of the entire vocal tract above it. Because the source and filter are assumed to be independent, it could also be referred to as an independent source and filter model. In **Figure 4A**, the major speech articulators, which can be divided into static and dynamic, are illustrated using a standard midsagittal view of the head. The tongue is a dynamic articulator, and speech is produced as a result of the interaction between the tongue with the static articulators, creating different tube configurations. **Figure 4B** illustrates the major static articulators in the approximate location they would fall on in a tube model of the vocal tract. For a neutral tube, the resonant frequencies can be calculated by assuming that the vocal tract is a closed-open tube and applying a one-quarter-length standing wave resonator to estimate the resonant frequencies of the tube. The tube model can be used to make predictions about the effect of different types of articulation and how they will impact the acoustic characteristics of the speech.

The Sounds of Language

As detailed by Ladefoged and Maddieson (1996), individual speech sounds, which are often referred to as segments or less accurately as phonemes, can be broken down into classes based on their production mechanisms and acoustic characteristics. The first main division is consonants and vowels. Vowels generally have a voicing source at the larynx (the structure that houses the vocal folds; **Figure 4A**), where egressive airflow (air flowing out) from the lungs sets the vocal folds in motion, and their spectral characteristics and resulting resonant, or *formant*, characteristics are determined by different vocal tract shapes above the larynx. The first three vocal tract resonances or formants (**Figure 5A**, *yellow bands*), together with the overall spectral shape of the signal, are the foundation for human perception of vowel quality (Hillenbrand et al., 2006). Vowels can also contrast in other ways. One way is in terms of duration, where they can vary in terms of long versus short vowels. Another way is whether the velopharyngeal port (the place where the velum and the pharyngeal wall meet; **Figure 4A**) is closed (with only oral resonances) or open (with additional nasal resonances due to airflow into the nasal cavity). Yet another way vowels contrast is whether they have a single main vowel quality (monophthongs) or vowel movement between two (diphthongs) or three (triphthongs) vowel qualities.

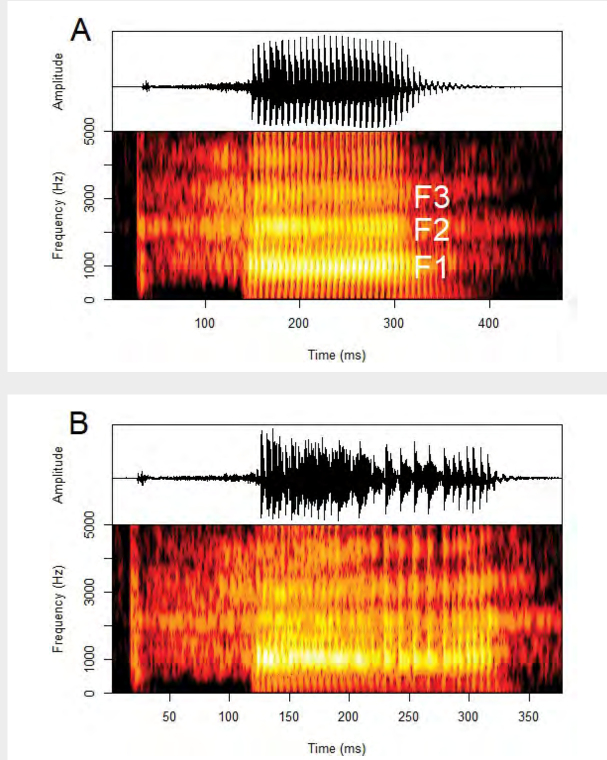


Figure 5. *Jalapa de Díaz Mazatec words with waveforms (**top**) and spectrograms (**bottom**) illustrating different voice qualities **A:** modal voicing [tʰæ], "itch" (word 21). **B:** creaky voicing [tʰæ̰], "sorcery" (word 20). F1, F2, and F3, first three vocal tract resonances or formants. Available at bit.ly/2TxrBdK from files bit.ly/2IjYG7H and bit.ly/32NE4OC produced by Speaker 4. Word numbers and speaker number reference the items in the original recordings.*

Voice Quality

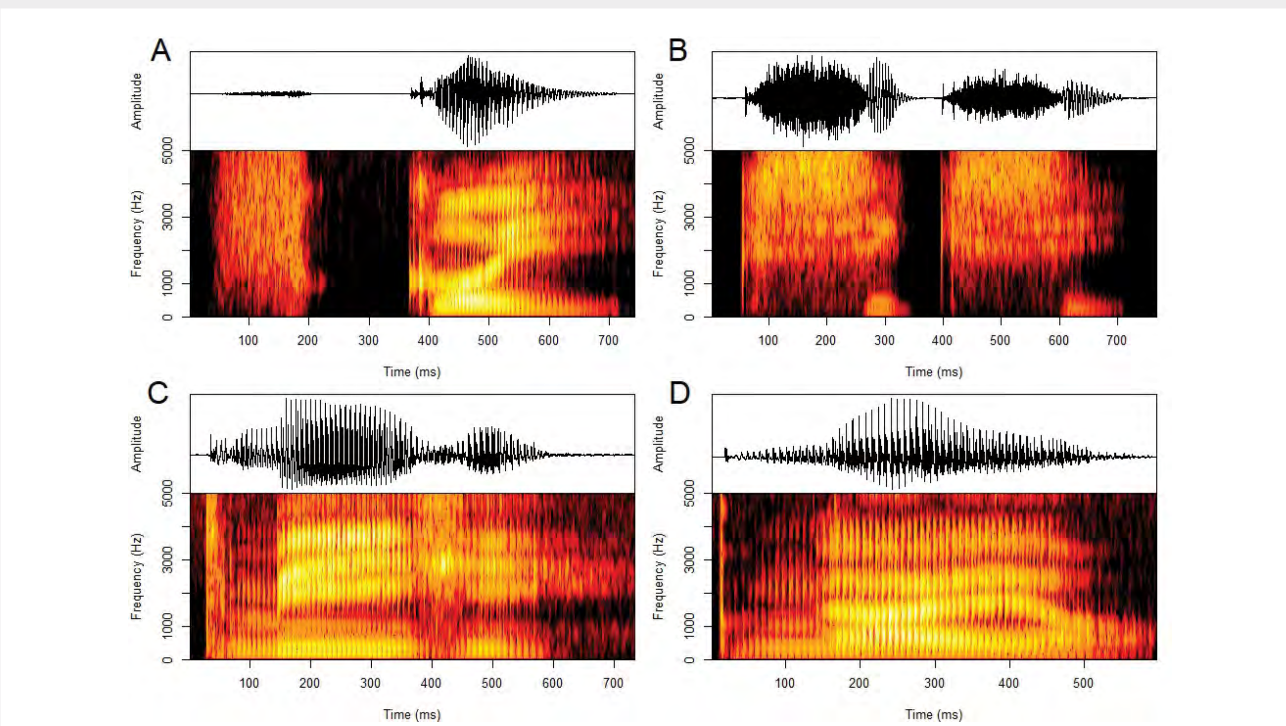
An important dimension of voiced segments are the ways in which speakers can manipulate vocal-fold vibrations creating distinct voice characteristics, referred to as voice quality or phonation type. Voice quality is typically classified into three types: modal, breathy, and creaky. Modal voicing is characterized by regular cycles and by a fairly linear drop in energy of about 6 dB/octave. Breathy voicing, in which the vocal folds are slack and very loosely held together, has a lower amplitude than modal voicing, and it is typified by an additional aperiodic component and a steep falloff in spectral energy. Creaky voicing, in which the vocal folds are slightly stiffer and tightly closed at the anterior end while allowing the posterior end to vibrate, has a lower amplitude than modal voicing, and it is characterized by longer, irregular, cycles and a shallow falloff in spectral energy.

Many languages employ differences in voice quality as a feature of lexical tone (the use of vocal pitch, for which the fundamental frequency is the acoustic correlate, to distinguish words), as in Mandarin and Vietnamese. For example, Mandarin has four main lexical tones (see **Multimedia1** at acousticstoday.org/tuckermedia): high level (as in the word "eight" 八 [pa˥]), mid-high rising (as in the word "to pull out" 拔 [pa˧˥]), mid-low-mid dipping (as in the word "to hold" 保 [pa˨˩˦]), and high-low falling (as in the word "father" 爸 [pa˥˩]). In the mid-low-mid dipping tone, creaky voicing is used. In other languages, fundamental frequency and voice quality can be used to convey meaning at the sentence level, as in English questions versus statements (see **Multimedia2** at acousticstoday.org/tuckermedia). Statements end in a low pitch that is often accompanied by creaky voicing.

Many other languages have contrastive (or phonemic) voice quality. Linguists use meaning differentiation to determine when speech sounds are contrastive in a language. For example, in English, *sit* and *zit* mean different things and are minimally contrastive; the sounds [s] and [z] are only distinguished by vocal-fold vibration, which is what differentiates the two words. One language that makes contrasts based on voice quality is Jalapa de Días Mazatec, an Otomanguean language spoken by about 17,500 speakers in Mexico (Eberhard et al., 2019). **Figure 5** illustrates two words where voice quality differences on the vowels conveys different meanings. In **Figure 5A**, [tʰæ], "itch" (see **Multimedia3** at acousticstoday.org/tuckermedia), the vocalic portion is modally voiced with regular cycles and a level amplitude. In **Figure 5B**, [tʰæ̰], "sorcery" (see **Multimedia4** at acousticstoday.org/tuckermedia), the vowel is realized with creaky (also known as *laryngealized*) voicing. The lower amplitude and longer and irregular cycles of creaky voicing can be seen between 225 and 300 ms in **Figure 5B**.

Some consonants, typically referred to as approximants, have dynamic vowel-like resonances, such as [w]. Like vowels, they are best defined in terms of their first three resonances (F1, F2, and F3). All other consonants can be described in terms of their place of articulation or where in the oral tract they are produced. These places of

Figure 6. *Waveforms (**top**) and spectrograms (**bottom**) illustrating complex consonant clusters in Tsou.* **A:** *[fkoi], "snake."* **B:** *[kʃikʃi], "ash/burning charcoal."* **C:** *[tmihi], "to hang."* **D:** *[pŋajo], "have food in mouth (actor focus)."*



articulation can be divided into the locations indicated in **Figure 4**. It is possible to have 17 places of articulation for consonants (Ladefoged and Maddieson, 1996). Consonants are also distinguished by the manner in which they are produced. For example, plosives are created by stopping the airflow in the oral tract and quickly releasing it as in the sound [p] in English. Fricatives are produced by making a tight constriction in the oral tract and creating turbulent airflow through the constriction, as in [s] in English. Laryngeal contrasts, as we have already seen, include voicing and voice quality and aspiration (a period of voicelessness following a plosive release). The airstream mechanism (how we control the flow of air in speech) is what a speaker manipulates to power speech and can be realized as pulmonic egressive, glottalic egressive/ejective, glottalic ingressive/implosive, or velaric ingressive/click. Egressive sounds are created by outward airflow; ingressive sounds are created using inward airflow.

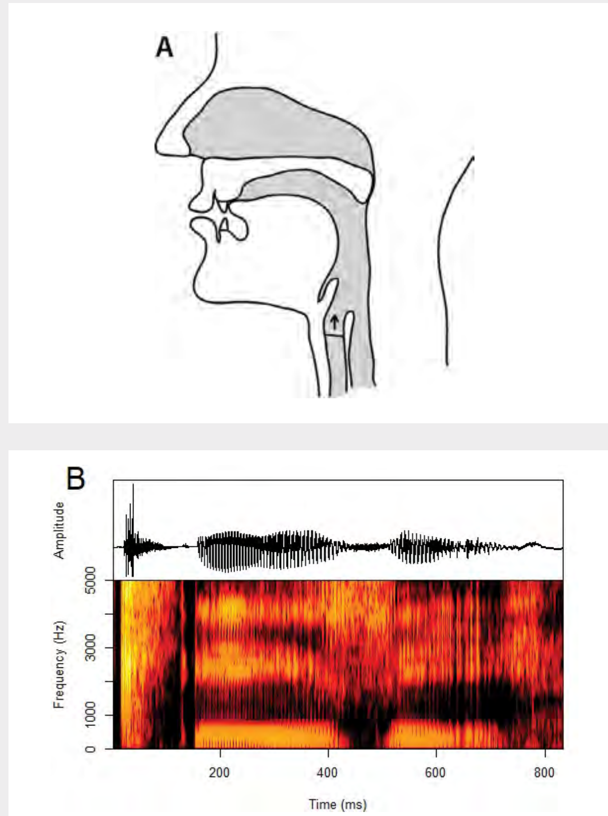


Figure 7. ***A:*** *midsagittal view of the speech articulators illustrating an alveolar ejective.* ***Arrow****, direction of laryngeal movement.* ***B:*** *waveform (****top****) and spectrogram (****bottom****) of [tł'i:ze], "a horse fly" in Dene Sųłiné.*

In addition to the segmental speech sounds, there are suprasegmental aspects of speech. A language's suprasegmental acoustic features include whether it is a tone language (e.g., Mandarin and Vietnamese, as seen in the Mandarin example), a stress language (e.g., English and Hawai'ian), or a pitch accent language (e.g., Japanese and Western Basque). The acoustic features of suprasegmental aspects of speech prominently include fundamental frequency as well as duration and intensity.

Phonotactics and Complex Combinations

A distinguishing characteristic of many languages is how they combine sounds, also called *phonotactics*. Although most languages have fairly simple phonotactics, some, like English, have much more complex combinations of segments as syllables as in the word *sports*, with a fricative ([s] in this case and plosive ([p] and [t]) combinations. A handful of known languages have very complex phonotactics. One example is Tsou, an Austronesian language spoken in Taiwan by about 4,000 speakers (Eberhard et al., 2019). Most of its consonants and vowels are commonly found in other languages. Unlike most languages, however, almost all of the two-way combinatorial possibilities of consonants are attested to in clusters at the beginning of words, resulting in many very rare combinations (Wright and Ladefoged, 1997). The examples (from recordings in Wright, 1996) in **Figure 6** illustrate combinations of [fk], [kʃ], [tm], and [pŋ] at the beginning of words. One of the interesting features in **Figure 6A** (see **Multimedia5** at acousticstoday.org/tuckermedia) with the fricative-stop cluster is the amplitude of the labiodental fricative. It has a much higher intensity, and it is longer than when it occurs in languages that only permit it to occur preceding a vowel. **Figure 6B** is interesting because of the extremely short vowels in comparison to the fricatives (see **Multimedia6** at acousticstoday.org/tuckermedia). **Figure 6**, **C** and **D**, is interesting in part because they have often been misperceived in impressionistic transcriptions as having an extra syllable (see **Multimedia7** and **8**, respectively, at acousticstoday.org/tuckermedia).

Nonpulmonic Airstream Mechanisms

Languages differ not only in how they use sounds but also in what sounds they have, how those sounds are made, and how they combine the sounds. For example, although all languages excite the vocal tract with pulmonic-egressive powered sources (at the larynx for vowels and at various points along the vocal tract for sounds with aperiodic sounds like fricatives, stops, and affricates), many languages use other sources, referred to as "airstream mechanisms."

Just over 10%, or 230, of the languages in the PHOIBLE sample contain ejective sounds (Moran and McCloy, 2019). Ejectives, illustrated in **Figure 7A**, are made using the laryngeal airstream mechanism as the source. It is first important to understand how these sounds are produced, and then we describe the acoustic characteristics of these sounds. Ejectives are made by first making a closure somewhere in the oral tract, like at the alveolar ridge (the hard ridge behind the upper teeth; **Figure 4A**). The speaker also closes the vocal folds and quickly raises the entire laryngeal system. The air trapped in the oral tract is compressed, increasing the air pressure in the oral tract. The tip of the tongue is lowered, opening the oral tract, and releasing the compressed air, creating an extreme popping sound. The waveform and spectrogram in **Figure 7B** are the same word recorded by Goddard in 1912 (**Figure 3**), [tɬʼi:ze], "a (horse) fly," by a speaker of Dene Sųłiné in 2020 (see **Multimedia9** at acousticstoday.org/tuckermedia). This word contains an ejectivized alveolar lateral affricate, which is an ejective that is released at the side of the tongue followed by a fricative. It can be seen in the spectrogram that the ejective release (between 0 and 50 ms) is the loudest part of the speech in the word, with strong transients in the waveform (e.g., Wright et al., 2002).

Figure 8. *Midsagittal view of the vocal tract illustrating the process of click production.* ***Dotted line****, closure both at the alveolar ridge with the tip of the tongue and at the velum with the back part of the tongue.* ***Arrow****, direction of the tongue movement to create the low-pressure area before release.*

Another airstream mechanism used is velaric. Sounds produced using this airstream mechanism are commonly known as clicks. By way of example, we describe the process of producing an alveolar click; this process is illustrated in **Figure 8**. First, the speaker raises the tongue and creates a closure both at the alveolar ridge with the tip of the tongue and at the velum with the back part of the tongue. The tongue is then pulled down while maintaining the closure at the alveolar ridge and velum. This creates an area of low pressure (a vacuum) within the cavity between the tongue and the roof of the mouth. Finally, the tip of the tongue is released from the alveolar ridge, creating a very loud popping sound as air fills the area of low pressure.

Clicks, like ejectives, are also extremely loud speech sounds. The high-intensity release of the click can clearly be seen in the four different realizations of the alveolar click in Ju|'hoan in **Figure 9**. These recordings come from the University of California, Los Angeles (UCLA) Phonetics Lab Archive dataset. Ju|'hoan, a Kx'a language, is spoken in Namibia and Botswana, with a population of about 44,000 speakers (Eberhard et al., 2019). Clicks represent just 1% or 29 of the languages spoken in the PHOIBLE sample (Moran and McCloy, 2019). One of the interesting things about clicks is that they can be combined with other sounds so that they are produced in many different ways. **Figure 9**, **A** and **B**, illustrates the plain alveolar click in the words [!ābē], "to be crinkled" (see **Multimedia10** at acousticstoday.org/tuckermedia), and [!āá], "to run" (see **Multimedia11** at acousticstoday.org/tuckermedia), respectively. The remaining examples illustrate the clicks occurring with different combinations of sounds. The example in **Figure 9C** illustrates a prenasalized alveolar click [ŋ!áā], "type of acacia" (see **Multimedia12** at acousticstoday.org/tuckermedia), and in **Figure 9D** is the voiced alveolar click from the word [ɡ!āà], "to dry something" (see **Multimedia13** at acousticstoday.org/tuckermedia).

As seen from the examples in this section, there is a wide variety of sounds in human language that are worthy of closer acoustic study. Some of the more interesting sounds are found in only a handful of languages, so one would need a very broad sample not to miss them.

Language Endangerment

We often hear about the tragedy of the loss of biological diversity, which is a major loss to the planet's ecosystem. Less often, we hear about cultural and linguistic diversity

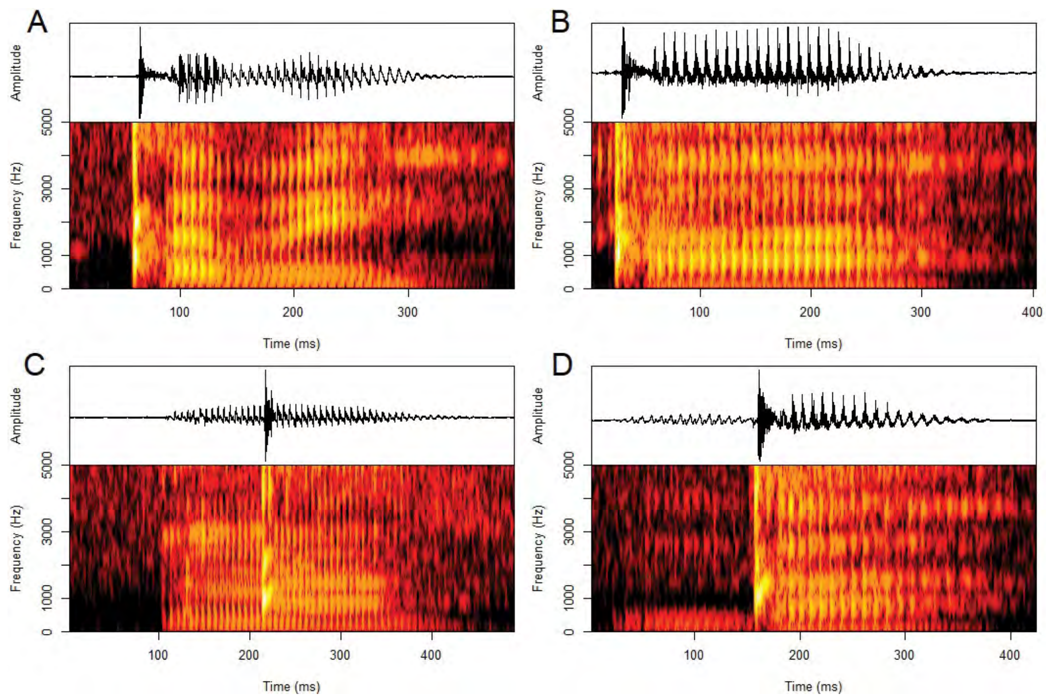


Figure 9. *Waveforms (**top**) and spectrograms (**bottom**) of Ju|'hoan words. **A:** plain alveolar click [!ābē], "to be crinkled" (word 3). **B:** plain alveolar click [!āá], "to run" (word 4). **C:** prenasalized alveolar click [ŋ!áā], "type of acacia" (word 7). **D:** voiced alveolar click [g!āà], "to dry something" (word 8). Available at bit.ly/2wvmaE8 from file bit.ly/2PNqGVn. Word numbers and speaker number reference the items in the original recordings.*

and how more than 75% of the world's languages have fewer than 1,000 speakers, and of those, many are in danger of losing all of their speakers within a generation. Many of these endangered languages exist in the most densely populated areas illustrated in **Figure 1B**. One resource calculates that over 40% of the world's languages are endangered (Eberhard et al., 2019). That is nearly 3,000 of the world's 7,000 languages that will likely not be spoken over the next 1-2 generations. Languages become endangered when future generations are not actively learning the community's language but are learning a more dominant language. The loss of language can play out in many different ways. An article in *Acoustics Today* by Whalen et al. (2011) gives an excellent example of some of the documentation of two endangered languages. The article also discusses in detail the importance of phonetic documentation from a language endangerment perspective.

Summary

As seen in the preceding examples, there is great diversity in the sounds of the world's languages, and much can be learned from these sounds. The investigation of the acoustic characteristics of the world's languages is an important part of understanding speech communication. Thus far, we have not mentioned the perception side of speech communication, and just as it is important to understand the acoustic characteristics of speech production across the world's languages, it is also important to understand how listeners of these languages make use of acoustic cues to comprehend language. We have already argued that the literature describing the acoustics of the world's languages is lacking; this lack of literature is even more extreme in the domain of speech perception for underdocumented languages. There are many speech sounds that vary radically from sounds in the well-documented languages, and most have not been studied acoustically. Similarly, a language may have a set of well-described sounds in its inventory but may combine them in a way that is not well documented; understanding how sounds interact with each and their acoustic characteristics is an important basic step to really understanding speech communication.

Acknowledgments

We thank Val Wood for providing the Dene Sųłiné example, Youran Lin for providing the Mandarin example, and Matthew Kelley for his comments on an early draft of this article.

References

Catford, J. C. (1977). *Fundamental Problems in Phonetics*. Indiana University Press, Bloomington.

Chiba, T., and Kajiyama, M. (1942). *The Vowel: Its Nature and Structure*. Tokyo-Kaiseikan Publishing Co., Ltd., Tokyo. Edition in reprinted in 1952; Japanese edition translated in 2003.

Eberhard, D. M., Simons, G. F., and Fennig, C. D. (Eds.). (2019). *Ethnologue: Languages of the World*, 22nd ed. SIL International, Dallas, TX.

Fant, G. (1960). Acoustic theory of speech production: With calculations based on X-ray studies of Russian articulations, vol. 2. *Description and Analysis of Contemporary Standard Russian*. Mouton, Hague, The Netherlands,. pp. 15-90.

Goddard, P. E. (1905). Mechanical aids to the study and recording of language. *American Anthropologist* 7(4), 613-619.

Goddard, P. E. (1912). Texts and analysis of Cold Lake dialect, Chipewyan. *Anthropological Papers of the American Museum of Natural History* 10(1-2), 170.

Hammarström, H., Forkel, R., and Haspelmath, M. (2019). *Glottolog 4.1*. Max Planck Institute for the Science of Human History, Jena, Germany. Available at http://glottolog.org. Accessed January 27, 2020.

Hillenbrand, J. M., Houde, R. A., and Gayvert, R. T. (2006). Speech perception based on spectral peaks versus spectral shape. *The Journal of the Acoustical Society of America* 119(6), 4041-4054. https://doi.org/10.1121/1.2188369.

International Phonetic Alphabet. (2018). International Phonetic Alphabet (IPA) and the IPA Chart. Available at http://www.internationalphoneticassociation.org/content/ipa-chart. Copyright © 2018 International Phonetic Association, available under a Creative Commons Attribution-Sharealike 3.0 Unported License.

Koenig, W., Dunn, H. K., and Lacy, L. Y. (1946). The sound spectrograph. *The Journal of the Acoustical Society of America* 18(1), 19-49. https://doi.org/10.1121/1.1916342.

Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages*. Blackwell, Oxford, UK.

Lindblom, B. (1990). On the notion of "possible speech sound." *Journal of Phonetics* 18, 135-152.

Lotz, J. (1950). Language and speech. *The Journal of the Acoustical Society of America* 22(6), 712-716. https://doi.org/10.1121/1.1906676.

Moran, S., and McCloy, D. (Eds.). (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History, Jena, Germany. Available at http://phoible.org. Accessed January 1, 2020.

Potter, R. K. (1945). Visible patterns of sound. *Science* 102(2654), 463-470. https://doi.org/10.1126/science.102.2654.463.

Proctor, M., Bresch, E., Byrd, D., Nayak, K., and Narayanan, S. (2013). Paralinguistic mechanisms of production in human "beatboxing": A real-time magnetic resonance imaging study. *The Journal of the Acoustical Society of America* 133(2), 1043-1054. https://doi.org/10.1121/1.4773865.

Rousselot, J.-P. (1897). *Principes de Phonétique Expérimentale*, tomes I and 2. Paris-Leipzig.

Traill, A. (1985). *Phonetic and Phonological Studies of !Xóõ Bushman*. Helmut Buske, Hamburg, Germany.

Tucker, B. V., and Wright, R. A. (2020). Introduction to the special issue on the phonetics of under-documented languages. *The Journal of the Acoustical Society of America* 147(4), 2741-2744. https://doi.org/10.1121/10.0001107.

UCLA Phonetics Lab. (2007). *The UCLA Phonetics Lab Archive*. Department of Linguistics, University of California Los Angeles (UCLA), Los Angeles, CA. Available at http://archive.phonetics.ucla.edu/.

Whalen, D. H., DiCanio, C., and Shaw, P. A. (2011). Phonetics of endangered languages. *Acoustics Today* 7(4), 35-42.

Wright, R. A. (1996). Tsou consonant clusters and cue preservation. UCLA PhD Dissertation, University of California, Los Angeles.

Wright, R. A., and Ladefoged, P. N. (1997). A phonetic study of Tsou. *Bulletin of the Institute of History and Philology*, Institute of History and Philology, Academia Sinica, Nanking, Taipei.

About the Authors

Benjamin V. Tucker
benjamin.tucker@ualberta.ca

Department of Linguistics
University of Alberta
4-32 Assiniboia Hall
Edmonton, Alberta T6G 2E7, Canada

Benjamin V. Tucker is a professor of phonetics in the Department of Linguistics, University of Alberta (Edmonton, AB, Canada). He is also the director of the Alberta Phonetics Laboratory. He received his PhD from the University of Arizona. His research focuses on cognitive aspects of the production and perception of spontaneous speech (e.g., "Wazat?" for "What is that?"). He also works on the documentation of endangered and underdocumented languages.

Richard Wright rawright@uw.edu

Department of Linguistics
University of Washington
Box 352425
Seattle, Washington 98195-2428, USA

Richard Wright is a professor of phonetics in the Department of Linguistics at the University of Washington (Seattle). He is also the director of the Linguistic Phonetics Laboratory. He received his PhD from the University of California, Los Angeles (UCLA; Los Angeles) in 1996. His research addresses systematic variation in the production and perception of spoken language. His other research interests include machine recognition of spoken language, hearing loss, and documentation of endangered and underdocumented languages.