# Phonetic Universals

Eleanor Chodroff

Department of Computational Linguistics, University of Zurich, Zurich, Switzerland; email: eleanor.chodroff@uzh.ch

**Keywords**

phonetic universals, phonetics–phonology interface, acoustic phonetics, dispersion, uniformity, articulatory ease

**Abstract**

Understanding the range and limits of crosslinguistic variation stands at the core of linguistic typology and more broadly, the scientific inquiry of human language. Linguistic typology is concerned with the relevant dimensions along which languages can vary and those along which they remain stable; an overarching goal is to understand the cognitive, physical, social, and historical factors that shape language. Phonetics is no exception to this enterprise, but it has faced obstacles in crosslinguistic data collection and processing power. The field has nevertheless established a solid foundation regarding the relevant dimensions of stability, revealing strong phonetic tendencies across languages (i.e., universals). This article provides an overview of phonetic universals with a summary of previously attested descriptive and analytic phonetic universals and consideration of methodological aspects when investigating phonetic universals. The increasing availability of multilingual speech data along with advanced speech processing tools promises a new era for investigations into crosslinguistic phonetic variation and systematicity.

## 1. INTRODUCTION

Understanding the range and limits of crosslinguistic variation stands at the core of linguistic typology and more broadly, the scientific inquiry of human language. Linguistic typology is concerned with the relevant dimensions along which languages can vary and those along which they remain stable; an overarching goal is to understand the cognitive, physical, social, and historical factors that shape language. As a working definition, a dimension along which languages demonstrate a high degree of stability can be termed a language universal. An understanding of universality in linguistics must go hand in hand with an understanding of language variation.

Historically, phonetics has been underrepresented in typological research. As detailed by Gordon (2016), only a small fraction of chapters in linguistic typology textbooks address even phonological typology, and none specifically addresses phonetic typology. This underrepresentation could stem from the controversial status of phonetics in linguistics (Chomsky & Halle 1968, p. 293). It could also reflect the challenge of inferring phonetic representations from acoustic or articulatory signals, which can require extensive crosslinguistic phonetic data and computational resources. However, recent technological advancements have facilitated greater access to such data and tools for phonetic analysis, laying the foundation for further exploration of phonetic typology. Fortunately, fundamental empirical and theoretical advances have already been made, even with limited access to truly crosslinguistic phonetic data.

This article provides an overview of phonetic universals.[1] It begins with a discussion of definitions, which is followed by a summary of previously attested descriptive and analytic phonetic universals and consideration of methodological aspects when investigating phonetic universals. Following the methodological considerations, I present a survey of highlighted empirical findings that relate descriptive phonetic universals to analytic phonetic universals.

## 2. DEFINING PHONETIC UNIVERSALS

### 2.1. Descriptive and Analytic Universals

At first blush, the term language universal brings along the implication of a property shared by all human languages—an absolute universal. Some argue that absolute universality can be accomplished using a large and comprehensive survey of existing languages (Piantadosi & Gibson 2014), though it can also be argued that full verification of an absolute universal would require access to data from all languages past and present. Another common understanding of a language universal is a linguistic property that occurs well above chance across the world's

[1]Given limitations on space, the article focuses on phonetic universals in the spoken modality, but some analytic factors could also be relevant for those in the signed modality (see, e.g., Sandler & Lillo-Martin, 2006).

languages. This latter definition is also referred to as a statistical universal or crosslinguistic tendency (Comrie 1989, Evans & Levinson 2009, Bickel 2015). Establishing statistical universals and crosslinguistic tendencies can be accomplished with access to sufficient, and sufficiently diverse, crosslinguistic data. This article uses the term universal in the sense of a statistical universal.

In discussing language universals, it can also be helpful to distinguish between empirical observations of crosslinguistic patterns (i.e., descriptive universals) and theoretical explanations of their origins (i.e., analytic universals) (Hyman 2008). Descriptive universals are empirical observations of highly consistent crosslinguistic phonetic patterns and highlight shared phonetic structures across languages; analytic universals offer explanations for why such surface phonetic variations occur and are theory dependent.

### 2.2. The Physiological Grounding of Phonetics

Unlike morphology or syntax, phonetics has a direct physiological grounding with constraints imposed by the human anatomy and motor system. In the discussion of universals in phonetics, it becomes pertinent to distinguish language or dialect variation from individual variation due to anatomy or biomechanics. Much of this discussion has also been divided into which aspects of variation should be called phonology, which aspects should be called phonetics, and whether both domains fall within the realm of the linguistic grammar. The discussion regarding the phonetics–phonology interface has been extensive (Cohn & Huffman 2014), and only a few primary points of consideration are raised here to highlight that many phonetic specifications are indeed language- or dialect-specific. Crosslinguistic patterns that arise over these might not necessarily be reduced to speech physiology; nevertheless, distinguishing physiological from alternative explanations of a crosslinguistic phonetic pattern is a constant theme in phonetic typology.

Early considerations of the phonetics–phonology interface posited that phonetic features are the discrete output of phonology and have two primary functions: a phonetic function and a classificatory function (Chomsky & Halle 1968). The features are universal and in their phonetic function potentially reflect several degrees of variation that reflect "independently controllable aspects of the speech event or independent elements of perceptual representation" and in their classificatory function are binary-valued (+ or –) and serve to represent the relevant phonological contrasts in the language (Chomsky & Halle 1968, p. 298). Discussion of the potential language-specific phonetic variation within the phonetic function of these features is minimal, and many potential aspects of variation, such as coarticulation, are shunted into universal phonetic rules (i.e., biomechanical constraints). It is unclear, then, if and how language-specific implementations that go beyond the classificatory function of a distinctive feature are part of the linguistic grammar. In other words, while a classificatory distinctive feature, such as [+voice], might suggest the presence of vocal fold vibration, the binary specification does not indicate potentially language-specific implementations such as the duration of vocal fold vibration, the relative timing of vocal fold vibration, or the rate of vocal fold vibration [e.g., fundamental frequency (f0), specified within the speaker's individual anatomical range].

Extensive evidence now exists, however, indicating that phonetic implementation, also referred to as phonetic realization, differs substantially across languages and dialects (Lisker & Abramson 1964, Disner 1983, Gordon et al. 2002, Fuchs & Toda 2010, Reidy 2016). For instance, the precise phonetic realization of a speech sound like [s] results in a higher peak frequency in English than in Japanese (Reidy 2016) and varies more generally from language to language (Gordon et al. 2002, Li et al. 2007, Fuchs & Toda 2010); it also varies by gender beyond any anatomical explanation (Heffernan 2004), sexual orientation (Linville 1998), and socioeconomic status (Stuart-Smith et al. 2003). These findings suggest that speakers exercise some control over the precise phonetic realization of a sound segment and that the variation cannot be wholly reduced to speech physiology. Importantly, there are several limitations of physiology in accounting for a wide variety of phonetic patterns (Keating 1985). Additional principles beyond biomechanical explanations are likely necessary to explain crosslinguistic phonetic tendencies.

Identifying the crosslinguistic patterns corresponds to an investigation of descriptive phonetic universals, and identifying the principles or constraints that account for that variation corresponds to an investigation of analytic phonetic universals. Alternative characterizations of phonetic universal include the contrast between mechanistic universals, which arise from automatic biomechanics of speech articulation, and ecological universals, which align with analytic factors like contrastivity and connectedness between speech sounds (Maddieson 1996). In the present article, the distinction between descriptive and analytic is preferred because the terminology is agnostic as to the source of an observed phonetic pattern. Indeed, much of the debate in phonetic typology concerns the distinction between automatic, biomechanical explanations of an observed phonetic pattern (Section 4.1)—that is, a mechanistic universal—and alternative explanations of an observed phonetic pattern.

## 3. DESCRIPTIVE UNIVERSALS

As discussed above, a descriptive phonetic universal denotes a consistent crosslinguistic phonetic pattern occurring above chance across languages. In phonetics, several such patterns have been identified; they are summarized in **Table 1** with a high-level overview and a relevant, early, but nonexhaustive set of references (for useful catalogs and descriptions, see also Keating 1984, Maddieson 1996). These are first presented here as a simple catalog. Many of these are discussed further in terms of their empirical support and analytic interpretation following the presentation of previously proposed analytic phonetic universals.

Table 1 A nonexhaustive list of putative descriptive phonetic universals that have been previously discussed in the literature

| **Phenomenon** | **Summary** | **Reference(s)** |
|---|---|---|
| Intrinsic vowel f0 | High vowels have a higher f0 than low vowels. | Keating 1985 |
| Intrinsic vowel duration | High vowels have a shorter duration than low vowels. | Keating 1985 |
| Extrinsic vowel duration | Vowels are shorter before or after a voiceless consonant than before or after a voiced consonant. | Maddieson 1996, Coretta 2019, cf. Keating 1985 |
| Vowel duration and syllable structure | Vowels in closed syllables (CVC) are shorter than vowels in open syllables (CV). | Maddieson 1996 |
| High vowel devoicing | High vowels are more susceptible to devoicing than low vowels. | Maddieson 1996 |
| Consonant f0 | Vowels following voiceless consonants have a higher f0 than vowels following voiced consonants. | Maddieson 1996 |
| Stop place of articulation and closure duration | Bilabial stops have longer closure duration than velar stops (or more posterior stops). | Maddieson 1996 |
| Stop place of articulation and | Bilabial stops have a shorter voice onset time than velar stops (or more posterior | Maddieson 1996 |

| | | |
|---|---|---|
| voice onset time | stops). | |
| Word-final devoicing | Voiceless stops are more likely than voiced stops to occur in word/utterance-final position. | Keating 1985 |
| Vowel-to-vowel coarticulation | Coarticulation from one vowel to another is greater in languages with smaller vowel inventories than in those with larger vowel inventories. | Manuel & Krakow 1984 |
| Nasal coarticulation | Vowels adjacent to a nasal consonant will also be partially nasalized, resulting in a forward or backward influence of the nasal. | Manuel & Krakow 1984 |
| Domain-initial strengthening | Segments are produced with more prominence or hyperarticulation at the beginning of a phonological phrase. | Fougeron 1998, Keating et al. 2004 |
| Phrase-final lengthening | Segments are longer toward the end of a phrase, particularly relative to their duration in phrase-medial position. | Maddieson 1996 |
| f0 declination and amplitude declination | f0 and amplitude decrease over the course of an utterance. | Maddieson 1996 |
| Rising f0 in polar questions | f0 rises in a polar (yes–no) question. | Ultan 1969, Bolinger 1978 |
| Deaccentuation of given information | Information that is given in a discourse has reduced prominence. | Cruttenden 2006 |

## 4. ANALYTIC UNIVERSALS

Analytic phonetic universals correspond to the explanatory principles that account for highly consistent crosslinguistic phonetic patterns—that is, descriptive phonetic universals. Identifying the factors that underlie phonetic variation and systematicity has been and will always be theory dependent. As mentioned in Section 2, the theoretical characterization of the relationship between phonetics and phonology varies considerably within the literature, and how exactly an author has conceptualized this relationship can complicate discussion of analytic phonetic universals. The differences can mostly be categorized into two primary perspectives: a direct relationship and an indirect relationship between phonetics and phonology (Gordon 2016). In a direct phonetic relationship, phonological units like phonemes correspond directly to phonetic realizations; discrete symbolic units are treated as substitutes for continuous phonetic variation. Conversely, an indirect phonetic relationship involves converting phonological units to phonetic representations, which may then be subject to additional universal or language-specific constraints. For clarity and comprehensiveness, this article generally assumes an indirect relationship as it allows for a thorough exploration of continuous phonetic variation across languages; however, many of the analytic universals discussed below were developed with the assumption of a direct relationship between phonetics and phonology.

This section employs four broad categories to organize the primary types of analytic phonetic universals proposed in the literature: automatic effects, contrast and dispersion, economy and uniformity, and ease.

### 4.1. Automatic Effects

Given its grounding in physiology, many crosslinguistic phonetic patterns may indeed stem from automatic physical effects of speech production. A common question in the investigation of descriptive phonetic universals is whether such effects are under speaker control or are merely by-products of speech articulation. For instance, pitch and amplitude declination during speech could result from decreased subglottal pressure over time after the initial breath. This may not be explicitly specified in the linguistic grammar; however, if evidence were found that the degree of pitch declination varied from language to language, it could arguably be under speaker control and thus specified in the grammar.

Similarly, intrinsic f0, the observation that high vowels have a higher f0 than low vowels, may be explained biomechanically via the tongue-pull hypothesis and associated jaw movement: tongue raising for high vowels might tighten the cricothyroid muscle, raising f0, akin to tightening a string on a guitar for higher pitch. Alternatively, lowering the jaw for low vowels could slacken the vocal folds, thereby lowering f0 (for an overview, see, e.g., Chen et al. 2021).

An implicit assumption in concluding automaticity is the notion that different speech sounds should—at some level—be specified in the same manner along a particular phonetic dimension (e.g., f0 should be the same for /i/ and /a/). Some might alternatively argue that the phonetic target is not explicitly defined but is rather underspecified for both sounds. Nevertheless, an observed difference in f0 between /i/ and /a/ would arise from automatic consequences of the anatomical and biomechanical constraints. Regardless of whether the phonetic dimension is explicitly specified in the grammar, the logic of a biomechanical explanation holds only when the assumption is for the observed dimension to have otherwise been the same across the two speech sounds. This assumption is explored further in Section 4.3 on economy and uniformity.

### 4.2. Contrast and Dispersion

Across the world's languages, the importance of contrast in sound systems is widely recognized. However, the mechanisms through which this phenomenon shapes linguistic systems have generated extensive debate. Various principles of contrast have been studied, particularly in relation to vowel inventories and, to a lesser extent, sibilant inventories. These principles aim to ensure that phonological segments are adequately spaced out in the phonetic space for perceptual distinctiveness. Principles of contrast might account for diverse effects, including broad effects on overall system organization and subtle factors like intrinsic vowel f0.

One of the foundational proposals is that vowel categories should be maximally dispersed within the relevant phonetic space (Jakobson 1941). Liljencrants & Lindblom (1972) introduced a quantitative model that defines vowels in terms of the formants F1, F2, and F3 in mel units but with F2 and F3 collapsed into a ratio, F2′. By minimizing the inverse distance between vowels, the model maximizes the overall distance of a given vowel inventory size within this two-dimensional space (F1 × F2′). Coordinates resulting from simulations for different inventory sizes were labeled with International Phonetic Alphabet (IPA) symbols based on their canonical formant values. While the resulting inventories closely resemble real-world observations, the model still has limitations. Compared to observed inventories, the model tends to exaggerate backness/frontness contrasts, underpredict schwa, and favor less frequent back unrounded vowels over more common front rounded ones. It also underpredicts symmetry: /o/ commonly co-occurs with /e/, but the model tends to favor /o/ paired with /ɛ/ given its greater phonetic distance from /o/. This issue is revisited in Section 4.3 on economy principles.

As an alternative to maximal contrast, Lindblom (1986) proposed that languages may instead settle for a principle of sufficient contrast, particularly in small vowel inventories. Building on this intuition, Adaptive Dispersion Theory predicts that within a given language, phonetic variation should be greater in a small vowel inventory than in a large vowel inventory. For example, in an /i a u/ system, the actual formants could occur anywhere around [i ɪ e] for /i/, [u o ʊ ɯ] for /u/, and [æ ɐ a ɑ ɒ] for /a/. Several empirical studies have followed up on this prediction but have shown inconclusive results (see Section 6.2).

Dispersion alone does not tell the whole story of phonetic variation in vowel systems. Beyond dispersion within a phonetic space, certain vowels are more common crosslinguistically due to their inherent properties. Quantal Theory (Stevens 1989) offers an explanation for the preference of some vowels over others, suggesting that certain acoustic regions are less affected by articulatory changes, leading to relative acoustic stability. Vowels in such stable regions may be more preferable across languages.

One common type of quantal space occurs when two or more formants within a vowel are close together. The widely observed three-vowel inventory /i a u/ is favored across languages even though /ɛ ɐ ʊ/ is equally dispersed. However, /i a u/ exhibit unique formant proximity: in /i/, F2 and F3 are nearly merged, while in /a/ and /u/, both F1 and F2 are close. Close formant proximity can create the perception of a single, merged formant (Chistovich & Lublinskaya 1979), allowing for greater articulatory freedom provided this single formant is achieved (Stevens 1989).

Building on Dispersion Theory and Quantal Theory, Schwartz et al. (1997a) presented a numerical implementation of Dispersion–Focalization Theory that incorporated both dispersion (Liljencrants & Lindblom 1972) and focalization (Stevens 1989) terms. Dispersion minimized inverse distance, while focalization prioritized segments with low intraformant distance, emphasizing acoustic and perceptual stability regions. The relative strengths of dispersion and focalization were adjustable via two parameters. Vowel inventory layouts were then predicted within an auditory formant-based space, operationalized as an F1 × F2′ Bark-scaled space, where F2′ incorporated F2, F3, and F4. After optimization, vowel labels were assigned based on the closest prototype vowel.

Dispersion–Focalization Theory offers the advantage of accommodating both extrinsic and intrinsic stability pressures (Abry et al. 1989; Schwartz et al. 1997a,b). By adjusting the strength of each constraint, the model predicts natural variation observed in vowel inventories worldwide. However, limitations still exist: The model still struggled to predict the prevalence of schwa and symmetrical vowel systems. Additionally, the model did not fully consider potential articulatory constraints (e.g., ease of articulation) that could influence vowel preferences.

Extending this approach**,** Cotterell & Eisner (2017, 2018) introduced a generative model for vowel inventories that not only addresses principles of dispersion and focalization but also addresses variation in inventory size. Apart from variation in vowel category locations within phonetic space, languages also vary in the number of vowel categories they possess. The model takes into consideration potential interactions between the overall number of vowels and their relative spacing.

Flemming's dissertation (Flemming 1995) and subsequent works implemented Dispersion Theory within an Optimality Theory framework. Three overarching constraints were proposed: (*a*) maximizing contrast distinctiveness, (*b*) minimizing articulatory effort, and (*c*) maximizing the number of contrasts (Flemming 1995, 2004). This theory balances competing constraints that govern phonological inventory structure, including perceptual distinctiveness, feature economy (see Section 4.3), and articulatory ease (see Section 4.4).

An additional concept in the realm of contrast and dispersion is feature enhancement (Kluender et al. 1988, Diehl & Kluender 1989, Kingston & Diehl 1994). It suggests that in the presence of a distinctive feature contrast (e.g., the [voice] difference between /p/ and /b/), speakers may use secondary phonetic dimensions like f0 or amplitude to boost perceptual contrast. This reinforcement enhances relevant auditory characteristics, resulting in a potential perceptual integration of the auditory dimensions, which may improve category recognition. Differences in vowel length between voiced and voiceless consonants, as well as consonant f0 effects, might be explained by such auditory motivations. With sufficiently large changes, these enhancements could also lead to sound change.

Finally, Dispersion Theory and quantal vowel regions alone have been argued to be insufficient to account for the phonetic patterns of crosslinguistic vowel systems. An alternative perspective is offered by Evolutionary Phonology, where vowel systems evolve across generations due to sound changes resulting from signal reanalysis prompted by factors like perceptual similarity, ambiguity, or choice (Vaux & Samuels 2015).

### 4.3. Economy and Uniformity

Another set of analytic phonetic universals focuses on principles of economy, uniformity, symmetry, and reuse of a phonetic target or gesture. These proposals vary in their assumptions regarding representations and the relationship between phonetics and phonology.

Maximal Utilization of Available Features (Ohala 1979) posits a direct relationship between phonetics and phonology. It suggests that languages should maximally use featural contrasts in their sound inventories, counteracting some undesirable predictions of dispersion. For instance, while dispersion might favor a mixed use of manners and places of articulation (e.g., [ɗ, k′, ts, ɬ, m, r, |]), languages typically opt for more symmetric and featurally economical systems. Similarly, Clements (2003a,b) proposed Feature Economy, stating that "languages tend to maximise the combinatory possibilities of features across the inventory of speech sounds"

(Clements 2003b: p. 287) and predicting, for example, that a language with /p t k/ is more likely to have /b d g/ than /ɖ ɟ ɢ/.

Maddieson's (1995) Gestural Economy involves a similar argument but at a phonetic level, suggesting that languages or individuals reuse physical gestures across segments. Gestures are considered physical and dynamic as well as distinct from abstract phonological features. The proposal also incorporates a principle of articulatorily efficient gestures that involve less extreme movements.

Additionally, the Maximal Utilization of Available Controls Theory (Schwartz et al. 2007, Ménard et al. 2008) implicates economy at an explicit substance-based, phonetic level. Building on Ohala's concept, the principle governs the use of controls, defined as "gestures shaped by multisensory perceptual mechanisms," that is, perceptuomotor targets, rather than abstract phonological features (Ménard et al. 2008: p. 15).

In a similar vein, Keating (2003) proposed constraints of articulatory and acoustic uniformity, where speakers prioritize near-identical articulation or acoustic realization across segments sharing a distinctive feature. Keating's study investigated the phonetic realization of the laryngeal feature in aspirated stop consonants across the place of articulation. Some speakers maintained a uniform glottal spreading gesture and timing relationship, while others exhibited near-identical voice onset times (VOTs) between /b/, /d/, and /g/. Speakers varied in whether uniformity operated on articulatory or acoustic levels, but the general concept enforces a dimension of similarity among distinct speech sounds.

Chodroff & Wilson (2017, 2022) proposed a target uniformity constraint that, rather than directly affecting articulatory or acoustic instantiation, promotes uniformity in the abstract phonetic targets that correspond to distinctive feature values. While a distinctive feature value provides a general idea of articulatory or auditory properties (e.g., [+anterior]), phonetic targets encode precise motor and auditory goals (e.g., tongue tip location); the mapping between them is referred to as phonetic realization. In a Bayesian model predicting acoustic correlates to phonetic targets, the target uniformity constraint is implemented as a prior distribution over secondary distinctive features that minimizes their influence. For instance, in a model predicting the acoustic correlate to sibilant place of articulation, the prior distribution over [voice] is centered on 0 with little variance, thus placing high prior probability over a lack of influence from the secondary feature. Although some deviation from perfect reuse of targets may occur, the constraint aims to minimize this relative to other factors like dispersion or articulatory ease.

Faytak (2018) has also argued for a critical role of uniformity in shaping the sound system of a language. The claim also has been made that this constraint arises from domain-general biases relating to articulation and articulatory reuse during acquisition (see also Faytak 2022).

Analogous to uniformity in phonetic realization, similar principles of uniformity may govern linguistic change and sociolinguistic phenomena. For instance, phonological categories with shared content often undergo parallel shifts in sound change (Fruehwald 2017), while in sociolinguistics, linguistic coherence may emerge from an economy principle (Guy & Hinskens 2016).

Furthermore, Chodroff & Wilson (2022) posited constraints of pattern uniformity and contrast uniformity that could contribute to the structure of phonetic inventories via conformity with the speaker population. Pattern uniformity promotes a consistent pattern of phonetic targets across speakers, enhancing population-level similarity in phonetic inventories. Contrast uniformity ensures a consistent difference between phonetic targets for opposing feature values. For instance, the distance between place of articulation targets for [s] and [ʃ], which contrast in [±anterior], should be uniform across speakers; this is a subcase of pattern uniformity and is limited to a featural contrast. These constraints differ from target uniformity in two key aspects: They enforce consistent differences rather than near-identity between phonetic targets, and they require comparisons across populations of speakers rather than within individual speakers.

### 4.4. Articulatory Ease

In addition to dispersion and uniformity, another constraint influencing phonetic realization is articulatory ease (Lindblom & Maddieson 1988, Lindblom 1990). Languages may prefer segments with simpler articulations and minimal effort (Boersma & Hamann 2008). Articulatory ease can in part be quantified by the number of gestures required to produce a segment (Lindblom & Maddieson 1988). Unlike economy constraints discussed above, which focus on reuse or uniformity of gestures within an inventory, articulatory ease pertains to the simplicity of articulation for individual speech sounds (Lindblom 1983, 1990). Thus, it differs from forms of articulatory reuse or uniformity. For instance, a language could have a complex set of articulations for a speech sound, but as long as this set is consistently reused across multiple sounds, the inventory remains economical, satisfying constraints like target uniformity. Related proposals of articulatory ease involve Lindblom's (1990) H&H Theory, which relates to a speaker's use of hypo- or hyperarticulation in speech production; in some cases, hypoarticulation may be easier to implement and sufficient for speech communication.

## 5. METHODOLOGICAL CONSIDERATIONS

Investigating phonetic universals requires several methodological considerations, including the dimensions along which languages will be compared (the meta-language), how crosslinguistic data are acquired, and whether the collected crosslinguistic data are sufficiently diverse for assessing a conclusion regarding universality.

### 5.1. The Meta-Language of Phonetics and Phonology

Establishing a set of generalizable units for comparison—that is, a meta-language—can offer significant advantages for identifying crosslinguistic phonetic patterns. A consistent meta-language enables direct language comparison and exploration of crosslinguistic variation. If we then look across a diverse set of languages using this unit of comparison, do we still observe a strong statistical generalization (Comrie 1989)? The choice of units in this meta-language has resulted in considerable debate: Should they reflect mental categories, should they be the most descriptively useful units, should they capture historical language change processes, and how does such standardization affect our understanding of language-specific nuances? Despite these discussions, a meta-language can be invaluable for defining language universals and exploring different types of universals.

In phonetics and phonology, meaningful units of speech have traditionally been represented by symbolic phonetic transcriptions such as IPA symbols, distinctive features, ToBI (tones and break indices) transcriptions, and semantic functions related to prosody. These symbols can have strong theoretical connotations, but they can also serve as standardized units that allow comparison across languages and facilitate extraction of acoustic or articulatory phonetic measurements.

Nevertheless, using established units for the meta-language of phonetics, such as IPA symbols, still comes with limitations. The description of languages with IPA symbols can vary considerably across researchers, with the type of variation ranging from what has previously been termed undernalysis to overanalysis (Anderson et al. 2023). Some researchers employ IPA symbols at a phonemic level to represent minimal pair contrasts only and potentially abstract over a wide range of phonetic realizations (underanalysis). This approach can lead to the loss of phonetic contrasts, which could hinder phonetic measurement and subsequent crosslinguistic comparison. Conversely, others use IPA symbols to faithfully represent phonetic details, often employing diacritics to account for minor variations (overanalysis). While this approach preserves finer phonetic distinctions, the fine-grained symbolic representation can complicate comparisons between different linguistic descriptions.

Similar issues will likely arise in the use of any discrete representational unit of language, including intonational units (e.g., ToBI or an alternative discrete system) and tone representations. Careful consideration of the chosen units and any theoretical commitment to their use is critical for engagement with this type of research. Regardless of the theoretical framework, these standardized abstractions—the meta-language—are valuable tools for comparing languages.

### 5.2. Data Collection

Crosslinguistic phonetic analyses have historically faced limitations in terms of the number of languages, speech sounds, and dimensions considered, partly due to computational constraints, data availability, and access to speech processing tools. Despite these challenges, several approaches have been established, including meta-analyses of existing data, laboratory-collected data analysis, and corpus analysis.

Meta-analyses involve aggregating standardized phonetic measurements from existing literature, ensuring comparability across studies and languages. While successful in investigating phonetic universals, this approach is restricted to a limited set of phonetic measurements that have been investigated in a consistent manner by various researchers. Notable meta-analyses include studies on vowel intrinsic f0 in 31 languages (Whalen & Levitt 1995), vowel F1 and F2 in over 200 languages (Becker-Kristal 2010), stop VOT in over 100 languages (Chodroff et al. 2019), and an examination of acoustic correlates of word stress in 75 language varieties (Gordon & Roettger 2017).

With some effort, phonetic universals can also be assessed through larger-scale laboratory data collection. Laboratory data offer the advantage of customized phonetic measurements applied consistently across languages with direct experimental control over potential confounds. However, laboratory studies have been severely limited in collecting large quantities of crosslinguistic data. Although online searches for "crosslinguistic phonetics" and "laboratory" yield many relevant studies, they typically involve only a handful of languages. The small sample size limits the overall generalizability of observed phonetic patterns to unseen languages.

The use of large-scale speech corpora has emerged as a promising avenue for investigating phonetic universals. Unlike laboratory data, corpus data are precollected for unrelated or more general purposes, but corpora can offer vast amounts of data for analysis. When appropriate statistical methods are applied, corpus data can prove highly conducive to a wide range of phonetic research questions. Similar to laboratory studies, this approach also allows for customized and consistent phonetic measurement across languages, but researchers are nevertheless limited by the availability of existing data and processing tools in this approach.

Corpus analyses have increased substantially in popularity, driven by advancements in computational power and the availability of crosslinguistic spoken data and speech processing tools. Publicly available crosslinguistic speech corpora include the UCLA Phonetics Lab Archive (Ladefoged et al. 2009), the CMU Wilderness Corpus (Black 2019), the Common Voice Corpus (Ardila et al. 2020), Multilingual LibriSpeech (Pratap et al. 2020), DoReCo (Paschen et al. 2020), and FLEURS (Conneau et al. 2023). Using speech processing tools like phonetic forced alignment and grapheme-to-phoneme (G2P) conversion, many of these corpora have been prepared for phonetic analysis with the inclusion of time-aligned phone-, word-, or phrase-level units [e.g., DoReCO (Paschen et al. 2020), VoxClamantis for Wilderness (Salesky et al. 2020), VoxCommunis for Common Voice (Ahn & Chodroff 2022), VoxAngeles for the UCLA Phonetics Lab Archive (Chodroff et al. 2024)]. Forced alignment tools for crosslinguistic data processing include the Montreal Forced Aligner (McAuliffe et al. 2017), WebMAUS (Kisler et al. 2017), and more recently, universal phone recognizers and aligners (Zhu et al. 2024). G2P resources include Epitran (Mortensen et al. 2018), WikiPron (Lee et al. 2020), the XPF Corpus (Cohen Priva et al. 2021), and CharsiuG2P (Zhu et al. 2022).

Example corpus phonetic studies in phonetic typology cover stop VOT in 18 languages (Cho & Ladefoged 1999), vowel formants in approximately 40 languages and sibilant spectral peak in 18 languages (Salesky et al. 2020), vowel formants in approximately 30 languages (Ahn & Chodroff 2022), vowel formants in 10 languages (Hutin & Allassonnière-Tang 2022), vowel f0 in 16 languages (Ting et al. 2024), and articulation rate in consonants and vowels across 8 typologically diverse languages (Lo & Sóskuthy 2023).

### 5.3. Sampling and Biases

As determining absolute universality presents an impossible task, researchers instead tend to rely on determining statistical universality. In a distributional sense, universality implies a prevalence of the phenomenon that is greater than chance across languages. Importantly, however, genealogical and areal biases in a given sample of languages must be controlled for to ensure that the effect is consistent across a diverse language sample. To obtain an unbiased representation of languages, several methods have been proposed. One is stratification, where the sample contains approximately equal and large numbers of language samples that are representative of their historical and geographic relationships. Another is to use more nuanced statistical methods that can control for nonindependence between observations such as hierarchical or mixed-effects regression models (for further discussion of language sampling biases and corrections, see Miestamo et al. 2016, Guzmán Naranjo & Becker 2022, Samardžić et al. 2024). Obtaining a diverse and statistically robust sample of languages remains a particularly challenging obstacle in phonetic typology.

Identifying a stratified and representative sample of languages can be enhanced through the use of typological resources, such as Grambank (Skirgård et al. 2023), Glottolog (Hammarström et al. 2024), and the World Atlas of Language Structures (WALS; Dryer & Haspelmath 2013). These resources contain an encyclopedia of linguist-determined phylogenetic relationships, macroareas, and grammatical features of each language.

## 6. EMPIRICAL INVESTIGATIONS OF PHONETIC UNIVERSALS

Many phonetic universals have been attributed to automatic, biomechanical factors, although some could also be explained by principles of dispersion or economy. This section presents a selection of empirical phonetic findings relating to automatic effects, dispersion, economy, and crosslinguistic suprasegmental features.

### 6.1. Empirical Investigations Relating to Automatic Effects

Many descriptive universals have been attributed to automatic, biomechanical consequences of speech production. An underlying assumption is that these automatic effects occur when the same implementation is used for a given phonetic dimension across two or more speech sounds. For example, while f0 may not be crucial for distinguishing /i/ from /a/, the tongue pulling on the larynx for /i/ may inadvertently raise f0 relative to /a/ (e.g., Fischer-Jørgensen 1990). The intention of a uniform implementation could reflect a principle of economy in phonetic inventories.

A critical debate in crosslinguistic differences, however, is whether certain phonetic perturbations are under speaker control or purely automatic. Speaker-controlled perturbations imply that phonetic targets are explicitly specified to produce the observed effect, which would allow for potential deliberate enhancement. Conversely, a purely biomechanical effect should yield consistent effect sizes across languages. However, variations in the magnitude of effects suggest some degree of speaker control. Keating (1985) argued that while biomechanics may explain the direction of these perturbations, the variability in magnitude across languages indicates the influence of other analytic factors. Alternative explanations beyond biomechanics are therefore necessary to fully account for these descriptive universals.

### 6.1.1. Intrinsic f0.

Intrinsic f0 refers to the observation that high vowels typically have higher f0 values than low vowels within a given language and speaker. Whalen & Levitt (1995) conducted a meta-analysis across 31 languages and 11 language families and confirmed the presence of this effect in high vowels ([i u ɯ]) versus low vowels ([a ɑ]). To investigate the influence of enhancement, they also explored the influence of vowel inventory size and found a slight but nonsignificant positive correlation. Moreover, this effect has even been observed in babbling among English- and French-acquiring infants, a finding that suggests an automatic effect rather than deliberate enhancement (Whalen et al. 1995). Thus, intrinsic f0 was considered a universal consequence of articulation and not subject to deliberate enhancement.

More recently, Ting et al. (2024) examined intrinsic f0 and consonant f0 across 16 languages from nine language families, with dozens to hundreds of speakers per language. The intrinsic f0 effect was observed in all languages but with significant differences in its strength and a smaller effect among tone languages. While acknowledging the potential articulatory basis of intrinsic f0, the authors suggested that the effect is likely still under speaker control and potentially modulated by vowel dispersion. An additional analysis also revealed a moderate positive correlation between the magnitude of the effect and vowel inventory size, indicating enhanced intrinsic f0 effects in languages with larger vowel inventories. Relatedly, Van Hoof & Verhoeven (2011) also identified a larger intrinsic f0 effect for Dutch (12-vowel inventory) than for Arabic (3-vowel inventory).

In addition, Chodroff et al. (2024) investigated intrinsic f0 between /i/ and /a/ across 53 languages from 17 language families and between /u/ and /a/ across 36 languages from 13 families. The expected direction was found in most but not all languages: Between /i/ and /a/, 74% of languages were consistent, and between /u/ and /a/, 89% were consistent. Though the study observed an overall lower conformance rate than previous crosslinguistic studies, each language was represented by only one speaker. In an investigation of four African tone languages, Connell (2002) also found conformity in only three of the four languages (consistent: Ibibio, Kunama, and Dschang; inconsistent: Mambila).

Additional studies have identified intrinsic f0 effects in individual languages, including American English (Shadle 1985), Angami and Mizo (Lalhminghlui et al. 2019), French and Italian (Kirby & Ladd 2016), Shona (Gonzales 2009), Taiwanese (Zee 1980), various English dialects (Jacewicz & Fox 2015), and Yoruba (Hombert 1977). The effect, however, can be modulated by various factors. For instance, among tone languages, the effect frequently disappears in low tones (Hombert 1977, Zee 1980, Whalen & Levitt 1995, Lalhminghlui et al. 2019); the effect is also smaller in nonprominent syllables (Ladd & Silverman 1984, Shadle 1985, Steele 1986) and lower pitch ranges (Ladd & Silverman 1984, Whalen & Levitt 1995).

**6.1.2. Intrinsic vowel duration.**

Intrinsic vowel duration refers to the observation that low vowels typically have longer durations than high vowels, and tense vowels typically have longer durations than lax vowels (House & Fairbanks 1953, Peterson & Lehiste 1960, Lindblom 1967, Keating 1985). This effect has been argued to reflect physical factors—namely, that the increased jaw displacement of low vowels requires a greater articulatory force, resulting in longer duration relative to high vowels (Lindblom 1967). The physical explanation has, however, been contested, and the effect may also be under speaker control with the potential for deliberate enhancement of the contrast (Westbury & Keating 1980, Solé & Ohala 2010).

This effect has been studied across various languages, including Catalan (Solé & Ohala 2010), Danish (Bundgaard 1980), English (House & Fairbanks 1953, Peterson & Lehiste 1960), Japanese (Solé & Ohala 2010), Swedish (Elert 1964, Lindblom 1967, Toivonen et al. 2015), and Thai (Abramson 1974). In a study of American English, Catalan, and Japanese, Solé & Ohala (2010) proposed a method for distinguishing automatic from controlled differences in vowel duration among high, mid, and low vowels. As speech rate increases, a stable vowel durational difference should indicate active control over the vowel-specific durational targets. Using this approach, they found that vowel duration is likely under speaker control for English and Catalan but is governed by mechanical phonetic factors in Japanese.

Toivonen et al. (2015) proposed that an automatic relationship between physical tongue height and vowel duration should result in a gradient relationship across individual vowel tokens. The correlation was examined between F1, representing tongue height, and vowel duration within each vowel category in English and Swedish. The correlation did not reach significance for any tested vowel qualities. Nevertheless, categorically high vowels showed longer durations on average than categorically low vowels. The lack of a trading relationship between tongue height and vowel duration adds further evidence against an automatic explanation for the observed effect.

#### 6.1.3. Consonant f0.

Consonant f0 refers to the tendency for vowels following phonologically voiceless consonants to have higher f0s compared to those following phonologically voiced consonants. This pattern remains consistent across various phonetic realizations of the laryngeal contrast (e.g., voiceless aspirated or unaspirated stops, phonetically voiced stops). The observed difference in f0 could potentially be attributed to automatic biomechanical factors in the implementation of phonetic voicing, assuming that f0 is intended to remain constant. The vertical larynx tension theory suggests that the lowering of the larynx during voiced obstruents helps sustain vocal fold vibration during closure. This results in easier voicing maintenance if the supraglottal pressure remains low, which can be achieved by enlarging the cavity (Hombert et al. 1979; see also Bell-Berti 1975, Westbury 1983, Maddieson 1984). Consequently, without any other alterations in implementation, a lowered larynx corresponds to a decreased f0.

This biomechanical explanation of consonant f0 would predict a decrease in f0 following voiced obstruents due to the lowering of the larynx during closure, making voicing easier to maintain. However, the consonant f0 effect is observed even after voiced and voiceless sonorants, where airflow is not obstructed, and voicing is relatively easier to maintain. For instance, in Burmese, voiced nasals and laterals contrast with voiceless counterparts, and the f0 difference is evident following these segments as well (Maddieson 1984). This suggests that there may be some degree of speaker control and potential enhancement involved in the phonetic contrast.

Perturbations in f0 following voiced versus voiceless consonants play a role in tonogenesis, the emergence of tone contrasts (Hombert et al. 1979). Indeed, consonant f0 is more prone to phonologization compared to intrinsic f0 even though both involve minor f0 contrasts. For instance, Seoul Korean has a sound change in progress involving consonant f0 and tonogenesis. This dialect has a three-way stop contrast (aspirated, lenis, and fortis stops) that was previously distinguished by VOT alone but now involves both VOT and f0 contrasts. Specifically, aspirated and lenis stops no longer differ in VOT, but they do differ in f0. Aspirated and lenis stops have a longer VOT than fortis stops, and aspirated stops have a higher onset f0 than lenis stops (Kang 2014). Covariation between tone and voicing is also observed in languages like Yabem (Austronesian) and Kammu (Mon-Khmer) (Kingston 2011). In Vietnamese, although covariation was initially present, the initial consonant voicing status was lost during the sound change.

With respect to empirical findings, Ting et al. (2024) observed a consistent direction in the consonant f0 effect in 16 investigated languages. As with intrinsic f0, however, the magnitude of the effect differed considerably across languages. Furthermore, the duration of the effect across the vowel can vary from language to language (see also Francis et al. 2006), and the overall effect can differ from speaker to speaker (Kirby et al. 2020, Pricop & Chodroff 2024). Additional empirical investigations of consonant f0 have been conducted in languages with a two-way true voicing contrast [Catalan (Pricop & Chodroff 2024), Dutch (Pinget & Quené 2023), French and Italian (Kirby & Ladd 2016), Spanish (Dmitrieva et al. 2015), Tokyo Japanese (Gao & Arai 2019)], a two-way aspirating contrast [American English (House & Fairbanks 1953, Lehiste & Peterson 1961, Hanson 2009), Cantonese (Francis et al. 2006, Luo 2018), German (Kohler 1982, Hoole & Honda 2011), Mandarin (Xu & Xu 2003, Luo 2018), Shanghai Chinese (Chen 2011), Swedish (Löfqvist 1975)], alternative two-way contrasts [Afrikaans (Coetzee et al. 2018), Swiss German (Ladd & Schmid 2018)], and three-way voicing contrasts [Khmer, Central Thai, and Vietnamese (Kirby 2018)].

### 6.1.4. Stop place of articulation and voice onset time.

For stop consonants that share the same laryngeal status, VOT shows an inverse relationship with place of articulation: Stops with more posterior places tend to have longer absolute VOTs (Fischer-Jørgensen 1954, Peterson & Lehiste 1960, Maddieson 1996b, Cho & Ladefoged 1999). This is particularly consistent for the ranking of labials and dorsals and has even been found in infant babbling (Whalen et al. 2007); however, the relative ranking of coronal stops tends to vary more across languages (Chodroff et al. 2019).

In a study of 18 languages from 12 language families, Cho & Ladefoged (1999) observed a consistently longer VOT in dorsal than in labial voiceless stops. In a meta-analysis of stop VOT from 147 language varieties and 36 language families, Chodroff et al. (2019) also observed a consistently longer VOT in dorsal than in labial stops with short-lag VOT (99% agreement). Among stops with long-lag and lead VOT, the rank was still consistent, but more variation was observed (long-lag: 84%; lead: 84%).

Cho & Ladefoged (1999) proposed several hypotheses to explain the increase in VOT with more posterior places of articulation. These hypotheses included aerodynamics principles, oral cavity size, articulatory movement and speed, extent of articulatory contact, glottal opening area change, and the temporal adjustment between stop closure duration and VOT, which necessitates a fixed duration of vocal fold opening. Among these, they found that only a fixed duration of vocal fold opening adequately explained the observed patterns in both aspirated and unaspirated stops.

The concept of a fixed timing relationship suggests an economy principle, where speakers may employ the same phonetic target for laryngeal features across various places of articulation (see also the "low-cost option" in Docherty 1992). However, Cho & Ladefoged (1999) acknowledged that languages might also have place-specific VOT targets for each stop. In their crosslinguistic analysis, highly predictable VOT relationships were observed across stops with the same laryngeal specification, but different places of articulation (a laryngeal series); moreover, the VOT differences between places were quite constrained. This predictability is consistent with a crosslinguistic tendency to maintain similar phonetic targets within a laryngeal series, aligning with principles of economy and uniformity.

### 6.2. Empirical Investigations Relating to Contrast and Dispersion

Several studies have explored the empirical implications of dispersion in vowel inventories and to a lesser extent in sibilant inventories. These investigations have primarily focused on two main predictions. First, phonetic segments should exhibit greater dispersion—meaning larger phonetic contrasts—within larger inventories compared to smaller ones. Second, according to Adaptive Dispersion Theory (Lindblom 1986), phonetic variability should also decrease as the inventory size decreases.

The findings regarding vowel inventories have been varied. In line with one concept of dispersion, larger formant frequency contrasts between point vowels have been observed in large inventories relative to small inventories. This pattern was observed in a study by Flege (1989) comparing English (large) to Spanish (small), in a study by Jongman et al. (1989) comparing German and English (large) to Greek (small), and in a study by Guion (2003) comparing Spanish (large) to Quichua (small). However, some studies have found no difference in formant frequency contrasts when examining peripheral vowels in large versus small inventories (Bradlow 1995, Meunier et al. 2003). In four dialects of Catalan with varying inventory sizes, Recasens & Espinosa (2009) found that smaller vowel systems were no less dispersed than larger ones, and there was no clear relation between the number of categories and overall variability.

Becker-Kristal (2010) conducted a meta-analysis of F1 and F2 means from over 300 languages and tested several predictions of dispersion. A reliable relationship was observed between the number of vowels and vowel space area. Moreover, an increase of peripheral vowels was correlated with a larger F1 range, and an increase of nonperipheral vowels was correlated with a larger F2 range. In addition, the phonetic realization of specific vowel categories was often found to be more variable across languages; the exact realization depended on the language-specific vowel inventory structure.

Another prediction of dispersion is an inverse relationship between inventory size and phonetic variability. This prediction has generally not been supported, except potentially among larger vowel inventories (Livijn 2000). For instance, in a study of 38 languages across 11 language families, Salesky et al. (2020) found no association between the number of vowel categories and a measure of variability. This measure assessed the joint conditional entropy of F1 and F2 given the vowel category, indicating how confusable the vowel categories were given an observed F1-F2 pairing. Similarly, Hutin & Allassonnière-Tang (2022) examined 10 languages and found no correlation between inventory size and the F1-F2 vowel area or between inventory size and F1 standard deviation (SD). Although a significant relationship was observed between inventory size and F2 SD, it contradicted the predicted direction.

The predictions from Dispersion Theory have also been explored in sibilant inventories. Empirical crosslinguistic analyses of fricative phonetics commonly found that spectral properties were well suited for many fricative contrasts (Nartey 1982, Gordon et al. 2002). Boersma & Hamann (2008) used the spectral mean [center of gravity (COG)] to simulate sound changes and predict when a language might acquire or lose a sibilant fricative.

In addition, Hauser (2022) employed COG to investigate dispersion effects in two-sibilant inventories (English, German) and three-sibilant inventories (Mandarin, Polish). Contrary to the expectations of Adaptive Dispersion Theory, no relationship was found between the number of sibilants in the inventory and COG variability. As an alternative, the proposed cue-weighting hypothesis suggests that dispersion might depend not only on a single phonetic dimension but also on the relative weighting of different dimensions in distinguishing sibilant contrasts. While COG might effectively differentiate /s/ from /ʂ/, another dimension like F2 might be more useful in distinguishing /s/ from /ɕ/. A comprehensive Dispersion Theory would thus need to consider the relative importance of each phonetic dimension for a given contrast.

**6.3. Empirical Investigations Relating to Economy and Uniformity**

Principles of economy and uniformity may explain the high similarity of a given phonetic dimension across otherwise contrastive speech sounds. In addition to the automatic effects discussed above, several studies have examined the predictions of an economy or uniformity constraint on phonetic realization.

Ménard et al. (2008) investigated the acoustic and articulatory stability of mid-high vowels (/e ø o/) and mid-low vowels (/ɛ œ ɔ/) in French. They identified a reuse of perceptuomotor targets for vowels of the same height, evidenced by stable F1 values across the F2 space and consistent tongue height across vowel pairs such as /e/ to /o/ and /ɛ/ to /ɔ/. This structural pattern, argued to be governed by the Maximal Utilization of Available Controls principle, suggests an economy of targets rather than dispersion. The authors argued that this structural pattern directly contradicts predictions of dispersion and instead reflects the Maximal Utilization of Available Controls, a principle of economy.

Similarly, several studies have found stability in F1 between front and back vowels with shared height specifications in various languages, including Philadelphia English (Fruehwald 2013), Yorkshire English (Watt 2000), dialects of Brazilian Portuguese (Oushiro 2019), American English, Canadian French, Continental French, Dutch, and Spanish (Schwartz & Ménard 2019), as well as crosslinguistically (Ahn & Chodroff 2022). In addition, Salesky et al. (2020) observed a strong correlation between language-specific midfrequency peaks of [s] and [z] across 18 languages from six language families, indicating an underlying identity or near-identity in the phonetic realization of the shared place of articulation feature.

### 6.4. Empirical Investigations Relating to Suprasegmental Patterns

Suprasegmental descriptive universals have also been investigated across languages, particularly regarding pitch patterns in different speech contexts. Bolinger (1978) conducted a crosslinguistic survey and found that terminal falls were predominant in statements (35 out of 37 languages), terminal rises were predominant in polar questions (37 out of 41 languages), and terminal falls were predominant in *wh*-questions (14 out of 17 languages). In a survey of 53 languages, Ultan (1969) similarly found rising terminals or high pitch in polar questions in almost all languages; the only exceptions were among languages with postpositions. However, the consistency of rising terminals or high pitch in naturalistic speech has been subject to debate (Geluykens 1988). Moreover, counterexamples have come to light: Belfast English and Chickasaw feature rising pitch in statements, while Roermond Dutch and Chickasaw exhibit falling pitch in questions. This variability suggests language specificity in intonational contours despite potentially universal relationships between the height and contours of f0 and their meaning (Ladd 1981, Ohala 1984).

Several empirical studies have also examined similarities and differences in the rhythmic profiles of languages. Traditionally, languages have been classified as either syllable-timed or stress-timed in terms of their rhythm, suggesting a universal dichotomy (Pike 1945; for an overview, see Grabe & Low 2002). While this classification may be overly simplistic (Bertinetto 1989, Arvaniti 2009), empirical evidence can shed light on the range of variation and potential patterns across languages. Ramus et al. (1999) conducted a study examining rhythm metrics in eight European languages and found a contrast between syllable and stress timing. In an analysis of 18 languages, Grabe & Low (2002) found an overall contrast between previously categorized stress- and syllable-timed languages but also a continuous range of rhythmic profiles. Rhythmic variation has also been investigated in Bulgarian, German, and Italian (Barry et al. 2003) and in Mandarin, Cantonese, and Thai (Dellwo et al. 2014), among other languages.

## 7. CONCLUSION

The increasing availability of multilingual speech data along with advanced speech processing tools presents a new era for investigations into crosslinguistic phonetic variation and systematicity. This endeavor necessitates a commitment to a meta-language for comparative analysis, though research communities may diverge in defining what precisely constitutes a phonetic universal. Regardless of the exact name, establishing empirical crosslinguistic phonetic patterns and identifying the corresponding analytic factors are critical to our understanding of phonetic diversity and phonetic typology more generally. With the rapid advances in computational power, crosslinguistic data availability, and speech processing tools, the phonetics community is well poised to examine phonetic patterns at scale. Phonetic theory and insight into the analytic factors underpinning phonetic structure can also grow from this strong empirical groundwork. After all, the strength of a theory is only as good as the quality of its supporting data.

## DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

Many thanks to Shayne Sloggett, Liang Zhao and an anonymous reviewer who provided helpful feedback. This work was supported by grant PR00P1_208460 from the Swiss National Science Foundation.

## LITERATURE CITED

Abramson A. 1974. Experimental phonetics in phonology: vowel duration in Thai. *Pasaa* 4(1):71–90

Abry C, Boë L-J, Schwartz J-L. 1989. Plateaus, catastrophes and the structuring of vowel systems. *J. Phonet.* 17(1–2):47–54

Ahn EP, Chodroff E. 2022. VoxCommunis: a corpus for cross-linguistic phonetic analysis. In *Proceedings of the 13th Language Resources and Evaluation Conference* (*LREC 2022*), ed. N Calzolari, F Béchet, P Blache, K Choukri, C Cieri, et al., pp. 5286–94. Marseille, Fr.: Eur. Lang. Res. Assoc.

Anderson C, Tresoldi T, Greenhill SJ, Forkel R, Gray R, List J-M. 2023. Variation in phoneme inventories: quantifying the problem and improving comparability. *J. Lang. Evol.* 8(2):149–68

Ardila R, Branson M, Davis K, Henretty M, Kohler M, et al. 2020. Common Voice: a massively-multilingual speech corpus. See Calzolari et al. 2020, pp. 4218–22

Arvaniti A. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66(1–2):46–63

Barry WJ, Andreeva B, Russo M, Dimitrova S, Kostadinova T. 2003. Do rhythm measures tell us anything about language type? See Solé et al. 2003, pp. 2693–96

Becker-Kristal R. 2010. *Acoustic typology of vowel inventories and dispersion theory: insights from a large cross-linguistic corpus.* PhD Diss., Univ. Calif., Los Angeles

Bell-Berti F. 1975. Control of pharyngeal cavity size for English voiced and voiceless stops. *J. Acoust. Soc. Am.* 57(2):456–61

Bertinetto PM. 1989. Reflections on the dichotomy 'stress' versus 'syllable-timing.' *Rev. Phonét. Appl*. 91(93):99–130

Bickel B. 2015. Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In *The Oxford Handbook of Linguistic Analysis*, ed. B Heine, H Narrog, pp. 901–23. Oxford, UK: Oxford Univ. Press. 2nd ed.

Black AW. 2019. CMU Wilderness Multilingual Speech Dataset. In *Proceedings of the 2019 International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), pp. 5971–75. Piscataway, NJ: IEEE

Boersma P, Hamann S. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25(2):217–70

Bolinger D. 1978. Yes–no questions are not alternative questions. In *Questions, Synthese Language Library*, Vol. 1, ed. H Hiż, pp. 87–105. Dordrecht, Neth.: Springer

Bradlow AR. 1995. A comparative acoustic study of English and Spanish vowels. *J. Acoust. Soc. Am.* 97(3):1916–24

Bundgaard M. 1980. An acoustic investigation of intrinsic vowel duration in Danish. *Annu. Rep. Inst. Phonet. Univ. Copenhagen* 14:99–119

Calzolari N, Béchet F, Blache P, Choukri K, Cieri C, et al., eds. 2020. *Proceedings of the 12th Language Resources and Evaluation Conference* (*LREC 2020*). Marseille, Fr.: Eur. Lang. Res. Assoc.

Chen W-R, Whalen DH, Tiede MK. 2021. A dual mechanism for intrinsic f0. *J. Phonet.* 87:101063

Chen Y. 2011. How does phonology guide phonetics in segment–f0 interaction? *J. Phonet.*

39(4):612–25
Chistovich LA, Lublinskaya VV. 1979. The “center of gravity” effect in vowel spectra and critical distance between the formants: psychoacoustical study of the perception of vowel-like stimuli. *Hear. Res*. 1:185–95
Cho T, Ladefoged P. 1999. Variation and universals in VOT: evidence from 18 languages. *J. Phonet.* 27(2):207–29
Chodroff E, Golden A, Wilson C. 2019. Covariation of stop voice onset time across languages: evidence for a universal constraint on phonetic realization. *J. Acoust. Soc. Am.* 145(1):EL109–15
Chodroff E, Pažon B, Baker A, Moran S. 2024. Phonetic segmentation of the UCLA Phonetics Lab Archive. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (*LREC-COLING 2024*), ed. N Calzolari, M-Y Kan, V Hoste, A Lenci, S. Sakti, N Xue, pp. 12724–33. Torino, Italy: Eur. Res. Lang. Assoc.
Chodroff E, Wilson C. 2017. Structure in talker-specific phonetic realization: covariation of stop consonant VOT in American English. *J. Phonet.* 61:30–47
Chodroff E, Wilson C. 2022. Uniformity in phonetic realization: evidence from sibilant place of articulation in American English. *Language* 98(2):250–89
Chomsky N, Halle M. 1968. *The Sound Pattern of English*. New York: Harper & Row
Clements GN. 2003a. Feature economy as a phonological universal. See Solé et al. 2003, pp. 371–74
Clements GN. 2003b. Feature economy in sound systems. *Phonology* 20(3):287–333
Coetzee AW, Beddor PS, Shedden K, Styler W, Wissing D. 2018. Plosive voicing in Afrikaans: differential cue weighting and tonogenesis. *J. Phonet.* 66:185–216
Cohen Priva U, Strand E, Yang S, Mizgerd W, Creighton A, et al. 2021. *The Cross-Linguistic Phonological Frequencies (XPF) Corpus*. Providence, RI: Dep. Cogn. Linguist. Psychol. Sci., Brown Univ.
Cohn AC, Huffman MK. 2014. Interface between phonology and phonetics. In *Oxford Bibliographies in Linguistics*. Oxford, UK: Oxford Univ. Press
Comrie B. 1989. *Language Universals and Linguistic Typology: Syntax and Morphology*. Chicago: Univ. Chicago Press

Conneau A, Ma M, Khanuja S, Zhang Y, Axelrod V, et al. 2023. FLEURS: few-shot learning evaluation of universal representations of speech. In *Proceedings of the 2022 IEEE Spoken Language Technology Workshop* (*SLT*), pp. 798–805. Piscataway, NJ: IEEE

Connell B. 2002. Tone languages and the universality of intrinsic $F0$: evidence from Africa. *J. Phonet.* 30(1):101–29

Coretta S. 2019. An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish. *Glossa* 4(1):125

Cotterell R, Eisner J. 2017. Probabilistic typology: deep generative models of vowel inventories. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vol. 1: *Long Papers*, ed. R Barzilay, M-Y Kan, pp. 1182–92. Stroudsburg, PA: Assoc. Comput. Linguist.

Cotterell R, Eisner J. 2018. A deep generative model of vowel formant typology. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1: *Long Papers*, ed. M Walker, H Ji, A Stent, pp. 37–46. Stroudsburg, PA: Assoc. Comput. Linguist.

Cruttenden A. 2006. The de-accenting of given information: a cognitive universal? *Empirical Approaches to Language Typology* 20(8):311–355

Dellwo V, Mok P, Jenny M. 2014. Rhythmic variability between some Asian languages: results from an automatic analysis of temporal characteristics. In *Proceedings of Interspeech 2014*, pp. 1708–11. Singapore: Int. Speech Commun. Assoc.

Diehl RL, Kluender KR. 1989. On the objects of speech perception. *Ecol. Psychol.* 1(2):121–44

Disner SF. 1983. *Vowel quality: the relation between universal and language-specific factors.* PhD Diss., Univ. Calif., Los Angeles

Dmitrieva O, Llanos F, Shultz AA, Francis AL. 2015. Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *J. Phonet.* 49:77–95

Docherty G. 1992. *The Timing of Voicing in British English Obstruents*. Berlin: De Gruyter

Dryer MS, Haspelmath M, eds. 2013. *WALS Online*. Leipzig, Ger.: Max Planck Inst. Evol. Anthropol. https://wals.info/

Elert C-C. 1964. *Phonologic studies of quantity in Swedish: based on material from Stockholm speakers*. PhD Diss., Univ. Stockholm, Stockholm

Evans N, Levinson SC. 2009. The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32(5):429–48

Faytak MD. 2018. *Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese.* PhD Diss., Univ. Calif., Berkeley

Faytak MD. 2022. Place uniformity and drift in the Suzhounese fricative and apical vowels. *Linguist. Vanguard* 8(s5):569–81

Fischer-Jørgensen E. 1954. Acoustic analysis of stop consonants. *Maître Phonét.* 32:42–59

Fischer-Jørgensen E. 1990. Intrinsic $F_0$ in tense and lax vowels with special reference to German. *Phonetica* 47(3–4):99–140

Flege JE. 1989. Differences in inventory size affect the location but not the precision of tongue positioning in vowel production. *Lang. Speech* 32(2):123–47

Flemming E. 1995. *Auditory representations in phonology.* PhD Diss., Univ. Calif., Los Angeles

Flemming E. 2004. Contrast and perceptual distinctiveness. In *Phonetically Based Phonology*, ed. B Hayes, R Kirchner, D Steriade, pp. 232–76. Cambridge, UK: Cambridge Univ. Press

Fougeron C. 1998. *Variations articulatoires en début de constituants prosodiques de différents niveaux en français.* PhD Diss., Univ. Paris 3–Sorbonne Nouv., Paris

Francis AL, Ciocca V, Wong VKM, Chan JKL. 2006. Is fundamental frequency a cue to aspiration in initial stops? *J. Acoust. Soc. Am.* 120(5):2884–95

Fruehwald J. 2013. *The phonological influence on phonetic change.* PhD Diss., Univ. Pa., Philadelphia

Fruehwald J. 2017. The role of phonology in phonetic change. *Annu. Rev. Linguist.* 3:25–42

Fuchs S, Toda M. 2010. Do differences in male versus female /s/ reflect biological or sociophonetic factors? In *Turbulent Sounds: An Interdisciplinary Guide*, ed. S Fuchs, M Toda, M Zygis, pp. 281–302. Berlin: De Gruyter Mouton

Gao J, Arai T. 2019. Plosive (de-)voicing and $f0$ perturbations in Tokyo Japanese: positional variation, cue enhancement, and contrast recovery. *J. Phonet.* 77:100932

Geluykens R. 1988. On the myth of rising intonation in polar questions. *J. Pragmat.* 12(4):467–85

Gonzales A. 2009. Intrinsic F0 in Shona vowels: a descriptive study. In *Selected Proceedings of the 39th Annual Conference on African Linguistics*, ed. A Ojo, L Moshi, pp. 145–55. Somerville, MA: Cascadilla

Gordon MK. 2016. *Phonological Typology*. Oxford, UK: Oxford Univ. Press

Gordon MK, Barthmaier P, Sands K. 2002. A cross-linguistic acoustic study of voiceless fricatives. *J. Int. Phonet. Assoc.* 32(2):141–74

Gordon MK, Roettger T. 2017. Acoustic correlates of word stress: a cross-linguistic survey. *Linguist. Vanguard* 3(1):20170007

Grabe E, Low EL. 2002. Durational variability in speech and the rhythm class hypothesis. In *Laboratory Phonology 7*, ed. C Gussenhoven, N Warner, pp. 515–46. Berlin/New York: De Gruyter Mouton

Guion SG. 2003. The vowel systems of Quichua-Spanish bilinguals: age of acquisition effects on the mutual influence of the first and second languages. *Phonetica* 602:98–128

Guy GR, Hinskens F. 2016. Linguistic coherence: systems, repertoires and speech communities. *Lingua* 172–173:1–9

Guzmán Naranjo M, Becker L. 2022. Statistical bias control in typology. *Linguist. Typol.* 26(3):605–70

Hammarström H, Forkel R, Haspelmath M, Bank S. 2024. *Glottolog 5.0*. Dataset, Max Planck Inst. Evol. Anthropol., Leipzig, Ger.

Hanson HM. 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *J. Acoust. Soc. Am.* 125(1):425–41

Hauser I. 2022. Speech sounds in larger inventories are not (necessarily) less variable. *J. Acoust. Soc. Am.* 152(5):2664–74

Heffernan K. 2004. Evidence from HNR that /s/ is a social marker of gender. *Toronto Work. Pap. Linguist*. 23:71–84

Hombert J-M. 1977. Consonant types, vowel height and tone in Yoruba. *Stud. Afr. Linguist.* 8(2):173–90

Hombert J-M, Ohala JJ, Ewan WG. 1979. Phonetic explanations for the development of tones. *Language* 55(1):37–58

Hoole P, Honda K. 2011. Automaticity versus feature-enhancement in the control of segmental F0. In *Where Do Phonological Features Come From?*, ed. GN Clements, R Ridouane, pp. 131–71. Amsterdam: John Benjamins

House AS, Fairbanks G. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J. Acoust. Soc. Am.* 25(1):105–13

Hutin M, Allassonnière-Tang M. 2022. Operation LiLi: using crowd-sourced data and automatic alignment to investigate the phonetics and phonology of less-resourced languages. *Languages* 7(3):234

Hyman LM. 2008. Universals in phonology. *Linguist. Rev.* 25(1–2):83–137

Jacewicz E, Fox RA. 2015. Intrinsic fundamental frequency of vowels is moderated by regional dialect. *J. Acoust. Soc. Am.* 138(4):EL405–10

Jakobson R. 1941. *Kindersprache, Aphasie und allgemeine Lautgesetze*. Frankfurt am Main, Ger.: Suhrkamp

Jongman A, Fourakis M, Sereno JA. 1989. The acoustic vowel space of Modern Greek and German. *Lang. Speech* 32(3):221–48

Kang Y. 2014. Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: a corpus study. *J. Phonet.* 45:76–90

Keating PA. 1985. Universal phonetics and the organization of grammars. In *Phonetic linguistics: essays in honor of Peter Ladefoged*, ed. V Fromkin, pp. 115–32. Orlando: Academic Press.

Keating PA. 2003. Phonetic and other influences on voicing contrasts. See Solé et al. 2003, pp. 20–23

Keating PA, Cho T, Fougeron C, Hsu C-S. 2004. Domain-initial articulatory strengthening in four languages. In *Phonetic Interpretation: Papers in Laboratory Phonology VI*, ed. J Local, R Ogden, R Temple, pp. 143–61. Cambridge, UK: Cambridge Univ. Press

Kingston J. 2011. Tonogenesis. In *The Blackwell Companion to Phonology*, ed. M Van Oostendorp, CJ Ewen, E Hume, K Rice. Malden, MA: Wiley-Blackwell

Kingston J, Diehl RL. 1994. Phonetic knowledge. *Language* 70(3):419–54

Kirby JP. 2018. Onset pitch perturbations and the cross-linguistic implementation of voicing: evidence from tonal and non-tonal languages. *J. Phonet.* 71:326–54

Kirby JP, Kleber F, Siddins J, Harrington J. 2020. Effects of prosodic prominence on obstruent-intrinsic F0 and VOT in German. In *Proceedings of Speech Prosody 2020*, ed. N Minematsu, M Kondo, T Arai, R Hayashi, pp. 210–14. Tokyo: Int. Speech Commun. Assoc.

Kirby JP, Ladd DR. 2016. Effects of obstruent voicing on vowel F0: evidence from “true voicing” languages. *J. Acoust. Soc. Am.* 140(4):2400–11

Kisler T, Reichel U, Schiel F. 2017. Multilingual processing of speech via web services. *Comput.*

*Speech Lang.* 45:326–47

Kluender KR, Diehl RL, Wright BA. 1988. Vowel-length differences before voiced and voiceless consonants: an auditory explanation. *J. Phonet.* 16(2):153–69

Kohler KJ. 1982. $F_0$ in the production of lenis and fortis plosives. *Phonetica* 39(4–5):199–218

Ladd DR. 1981. On intonational universals. In *Advances in Psychology*, Vol. 7: *The Cognitive Representation of Speech*, ed. T Myers, J Laver, J Anderson, pp. 389–97. Amsterdam: North-Holland

Ladd DR, Schmid S. 2018. Obstruent voicing effects on F0, but without voicing: phonetic correlates of Swiss German lenis, fortis, and aspirated stops. *J. Phonet.* 71:229–48

Ladd DR, Silverman KEA. 1984. Vowel intrinsic pitch in connected speech. *Phonetica* 41(1):31–40

Ladefoged P, Blankenship B, Schuh RG, Jones P, Gfroerer N, et al. 2009. The UCLA Phonetics Lab Archive. *UCLA Department of Linguistics*. http://archive.phonetics.ucla.edu

Lalhminghlui W, Terhiija V, Sarmah P. 2019. Vowel-tone interaction in two Tibeto-Burman languages. In *Proceedings of Interspeech 2019*, pp. 3970–74. Graz, Austria: Int. Speech Commun. Assoc.

Lee JL, Ashby LFE, Garza ME, Lee-Sikka Y, Miller S, et al. 2020. Massively multilingual pronunciation modeling with WikiPron. See Calzolari et al. 2020, pp. 4223–28

Lehiste I, Peterson GE. 1961. Some basic considerations in the analysis of intonation. *J. Acoust. Soc. Am.* 33(4):419–25

Li F, Edwards J, Beckman M. 2007. Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. In *Proceedings of the 16th International Congress of Phonetic Sciences*, ed. J Trouvain, WJ Barry, pp. 917–20. Saarbrücken, Ger.: Univ. Saarl.

Liljencrants J, Lindblom B. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48(4):839–62

Lindblom B. 1967. Vowel duration and a model of lip mandible coordination. *Speech Transm. Lab. Q. Prog. Status Rep.* 4:1–29

Lindblom B. 1983. Economy of speech gestures. In *The Production of Speech*, ed. PF MacNeilage, pp. 217–45. New York: Springer

Lindblom B. 1986. Phonetic universals in vowel systems. In *Experimental Phonology*, ed. JJ Ohala, J Jaeger, pp. 13–44. Orlando: Academic

Lindblom B. 1990. Explaining phonetic variation: a sketch of the H&H Theory. In *Speech Production and Speech Modelling*, ed. WJ Hardcastle, A Marchal, pp. 403–39. Dordrecht, Neth.: Springer

Lindblom B, Maddieson I. 1988. Phonetic universals in consonant systems. In *Language, Speech, and Mind*, ed. LM Hyman, C Li, pp. 62–78. London/New York: Routledge

Linville SE. 1998. Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia Phoniatr. Logopaed.* 50(1):35–48

Lisker L, Abramson AS. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20(3):384–422

Livijn P. 2000. Acoustic distribution of vowels in differently sized inventories—hot spots or adaptive dispersion? In *Proceedings of the 13th Swedish Phonetics Conference (FONETIK 2000),* pp. 93–96, Skövde, Sweden: Univ. Stockholm.

Lo RY-H, Sóskuthy M. 2023. Articulation rate in consonants and vowels: results and methodological challenges from a cross-linguistic corpus study. In *Proceedings of the 20th International Congress of Phonetic Sciences*, ed. R Skarnitzl, J Volín, pp. 3206–10. Prague: Int. Phonet. Assoc.

Löfqvist A. 1975. Intrinsic and extrinsic $F_o$ variations in Swedish tonal accents. *Phonetica* 31(3–4):228–47

Luo Q. 2018. *Consonantal effects on F0 in tonal languages*. PhD Diss., Mich. State Univ., East Lansing

Maddieson I. 1984. The effects on $F_0$ of a voicing distinction in sonorants and their implications for a theory of tonogenesis. *J. Phonet.* 12(1):9–15

Maddieson I. 1995. Gestural economy. In *Proceedings of the 13th International Congress of Phonetic Sciences,* ed K Elenius, P Branderud, pp. 574–577. Stockholm

Maddieson I. 1996. Phonetic universals. In *UCLA Work. Pap. Phonet.* 92, pp. 160–78

Manuel SY, Krakow RA. 1984. Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Lab. Status Rep. Speech Res.* 77(78):69–78

McAuliffe M, Socolof M, Mihuc S, Wagner M, Sonderegger M. 2017. Montreal Forced Aligner: trainable text-speech alignment using Kaldi. In *Proceedings of Interspeech 2017*, pp. 498–502. Stockholm: Int. Speech Commun. Assoc.

Ménard L, Schwartz J-L, Aubin J. 2008. Invariance and variability in the production of the height feature in French vowels. *Speech Commun*. 50(1):14–28

Meunier C, Frenck-Mestre C, Lelekov-Boissard T, Le Besnerais M. 2003. Production and perception of vowels: Does the density of the system play a role? See Solé et al. 2003, pp. 723–26

Miestamo M, Bakker D, Arppe A. 2016. Sampling for variety. *Linguist. Typol.* 20(2):233–96

Mortensen DR, Dalmia S, Littell P. 2018. Epitran: precision G2P for many languages. In *Proceedings of the 11th International Conference on Language Resources and Evaluation* (*LREC 2018*), pp. 2710–14. Miyazaki, Jpn.: Eur. Res. Lang. Assoc.

Nartey JNA. 1982. *On fricative phones and phonemes: measuring the phonetic differences within and between languages*. PhD Diss., Univ. Calif., Los Angeles

Ohala JJ. 1979. The contribution of acoustic phonetics to phonology. In *Frontiers of Speech Communication Research*, ed. B Lindblom, S Öhman, pp. 355–63. London: Academic

Ohala JJ. 1984. An ethological perspective on common cross-language utilization of $F_0$ of voice. *Phonetica* 41(1):1–16

Oushiro L. 2019. Linguistic uniformity in the speech of Brazilian internal migrants in a dialect contact situation. In *Proceedings of the 19th International Congress of Phonetic Sciences*, ed. S Calhoun, P Escudero, M Tabain, P Warren, pp. 686–90. Canberra: Aust. Speech Sci. Technol. Assoc.

Paschen L, Delafontaine F, Draxler C, Fuchs S, Stave M, Seifart F. 2020. Building a time-aligned cross-linguistic reference corpus from language documentation data (DoReCo). See Calzolari et al. 2020, pp. 2657–66

Peterson GE, Lehiste I. 1960. Duration of syllable nuclei in English. *J. Acoust. Soc. Am.* 32(6):693–703

Piantadosi ST, Gibson E. 2014. Quantitative standards for absolute linguistic universals. *Cogn. Sci.* 38(4):736–56

Pike KL. 1945. *The Intonation of American English*. Ann Arbor: Univ. Mich. Press

Pinget A-F, Quené H. 2023. Effects of obstruent voicing on vowel fundamental frequency in Dutch. *J. Acoust. Soc. Am.* 154(4):2124–36

Pratap V, Xu Q, Sriram A, Synnaeve G, Collobert R. 2020. MLS: a large-scale multilingual dataset for speech research. In *Proceedings of Interspeech 2020*, pp. 2757–61. Shanghai: Int. Speech Commun. Assoc.

Pricop B, Chodroff E. 2024. Consonant f0 effects: a case study on Catalan. In *Proceedings of Speech Prosody 2024*, pp. 329–33. Leiden, Neth.: Int. Speech Commun. Assoc.

Ramus F, Nespor M, Mehler J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3):265–92

Recasens D, Espinosa A. 2009. Dispersion and variability in Catalan five and six peripheral vowel systems. *Speech Commun.* 51(3):240–58

Reidy PF. 2016. Spectral dynamics of sibilant fricatives are contrastive and language specific. *J. Acoust. Soc. Am.* 140(4):2518–29

Salesky E, Chodroff E, Pimentel T, Wiesner M, Cotterell R, et al. 2020. A corpus for large-scale phonetic typology. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ed. D Jurafsky, J Chai, N Schluter, J Tetrault, pp. 4526–46. Online: Assoc. Comput. Linguist.

Sandler W, Lillo-Martin DC. 2006. *Sign Language and Linguistic Universals*. Cambridge, UK: Cambridge Univ. Press

Samardžić T, Gutierrez X, Bentz C, Moran S, Pelloni O. 2024. A measure for transparent comparison of linguistic diversity in multilingual NLP data sets. In *Findings of the Association for Computational Linguistics: NAACL 2024*, ed. K Duh, H Gomez, S Bethard, pp. 3367–3382, Mexico City, Mex. Assoc. Comput. Linguist.

Schwartz J-L, Boë L-J, Abry C. 2007. Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAF) principle in a Perception-for-Action-Control Theory (PACT). In *Experimental Approaches to Phonology*, ed. MJ Solé, PS Beddor, M Ohala, pp. 104–24. Oxford, UK: Oxford Univ. Press

Schwartz J-L, Boë L-J, Vallée N, Abry C. 1997a. The Dispersion-Focalization Theory of vowel systems. *J. Phonet.* 25(3):255–86

Schwartz J-L, Boë L-J, Vallée N, Abry C. 1997b. Major trends in vowel system inventories. *J. Phonet.* 25(3):233–53

Schwartz J-L, Ménard L. 2019. Structured idiosyncrasies in vowel systems. *OSF Preprints*. https://osf.io/preprints/b6rdv/

Shadle CH. 1985. Intrinsic fundamental frequency of vowels in sentence context. *J. Acoust. Soc. Am.* 78(5):1562–67

Skirgård H, Haynie HJ, Blasi DE, Hammarström H, Collins J, et al. 2023. Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Sci. Adv.* 9(16):eadg6175

Solé M-J, Ohala JJ. 2010. What is and what is not under the control of the speaker: intrinsic vowel duration. In *Laboratory Phonology 10*, ed. C Fougeron, B Kühnert, M D'Imperio, N Vallée, pp. 607–56. Berlin/New York: De Gruyter Mouton

Solé M-J, Recasens D, Romero J, eds. 2003. *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Int. Phonet. Assoc.

Steele SA. 1986. Nuclear accent *F*0 peak location: effects of rate, vowel, and number of following syllables. *J. Acoust. Soc. Am.* 80(S1):S51

Stevens KN. 1989. On the quantal nature of speech. *J. Phonet.* 17:3–45

Stuart-Smith J, Timmins C, Wrench A. 2003. Sex and gender differences in Glaswegian /s/. See Solé et al. 2003, pp. 1851–54

Ting C, Clayards M, Sonderegger M, McAuliffe M. 2024. The cross-linguistic distribution of vowel and consonant intrinsic F0 effects. *Language*. In press

Toivonen I, Blumenfeld L, Gormley A, Hoiting L, Logan J, et al. 2015. Vowel height and duration. In *Proceedings of the 32nd West Coast Conference on Formal Linguistics*, ed. U Steindl, T Borer, H Fang, A García Pardo, P Guekguezian, et al., pp. 64–71. Somerville, MA: Cascadilla

Ultan R. 1969. Some general characteristics of interrogative systems. *Work. Pap. Lang. Univers.* 1:40–63

Van Hoof S, Verhoeven J. 2011. Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *J. Phonet.* 39(2):168–77

Vaux B, Samuels B. 2015. Explaining vowel systems: dispersion theory versus natural selection. *Linguist. Rev.* 32(3):573–99

Watt DJL. 2000. Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated? *Lang. Var. Change* 12(1):69–101

Westbury JR. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *J. Acoust. Soc. Am.* 73(4):1322–36

Westbury JR, Keating PA. 1980. Central representation of vowel duration. *J. Acoust. Soc. Am.* 67(S1):S37

Whalen DH, Levitt AG. 1995. The universality of intrinsic $F_0$ of vowels. *J. Phonet.* 23(3):349–

66
Whalen DH, Levitt AG, Goldstein LM. 2007. VOT in the babbling of French- and English-learning infants. *J. Phonet.* 35(3):341–52
Whalen DH, Levitt AG, Hsiao P-L, Smorodinsky I. 1995. Intrinsic *F*0 of vowels in the babbling of 6-, 9-, and 12-month-old French- and English-learning infants. *J. Acoust. Soc. Am.* 97(4):2533–39
Xu CX, Xu Y. 2003. Effects of consonant aspiration on Mandarin tones. *J. Int. Phonet. Assoc.* 33(2):165–81
Zee E. 1980. Tone and vowel quality. *J. Phonet.* 8(3):247–58
Zhu J, Yang C, Samir F, Islam J. 2024. The taste of IPA: towards open-vocabulary keyword spotting and forced alignment in any language. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1: *Long Papers*, ed. K Duh, H Gomez, S Bethard, pp. 750–72. Mexico City: Stroudsburg, PA: Assoc. Comput. Linguist.
Zhu J, Zhang C, Jurgens D. 2022. ByT5 model for massively multilingual grapheme-to-phoneme conversion. In *Proceedings of Interspeech 2022*, pp. 446–50. Incheon, South Korea: Int. Speech Commun. Assoc.