Phonetic Universals

Eleanor Chodroff
University of Zurich

Abstract: Understanding the range and limits of crosslinguistic variation stands at the core of linguistic typology and basic science. Linguistic typology is concerned with the relevant dimensions along which languages can vary, and in complement to that, the relevant dimensions along which languages are stable. Phonetics is no exception to this enterprise, and considerable advances have been made in developing theoretical and empirical findings within the field of phonetic typology. The present article provides an overview to phonetic universals: how should a phonetic universal be defined, how might a phonetic universal be investigated, and what is the current state of knowledge with respect to descriptive phonetic universals, i.e., empirically attested phonetic universals and analytic phonetic universals, i.e., principles or constraints that might account for empirical phonetic patterns.

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

1. Introduction

Understanding the range and limits of crosslinguistic variation stands at the core of linguistic typology and scientific inquiry. Linguistic typology explores both the dimensions along which languages can vary and those along which they remain stable. An understanding of universality in linguistics must go hand-in-hand with an understanding of language variation.

Historically, phonetics and phonology have been underrepresented in typological research. Gordon (2016) notes that few chapters in linguistic typology textbooks address phonological typology, with none specifically addressing 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 acoustic-phonetic analysis, laying the foundation for further exploration of phonetic typology. Fortunately, fundamental empirical and theoretical advances have already been made for phonetic typology, even with limited access to truly crosslinguistic phonetic data.

The discussion of phonetic typology, particularly as distinct from phonological typology, is further complicated by varying characterization of the relationship between phonetics and phonology. Some research that falls under phonological typology might also be relevant for phonetic typology. This distinction often hinges on whether the relationship between phonology and phonetics is assumed to be direct or indirect.

In a direct phonetic relationship, phonological units like phonemes directly correspond to phonetic realizations, treating discrete symbolic units 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. The direct phonetic relationship tends to operate on discrete symbolic units as a direct substitute for continuous phonetic variation.

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, we acknowledge the relevance of proposals based on a direct phonetic relationship, and we will occasionally reference this distinction in our discussion of phonetic typology.

The chapter focuses on identifying universal aspects of phonetics and understanding consistent crosslinguistic patterns in phonetic realization. It comprises several sections: defining phonetic universals, methodological approaches to investigating phonetic universals, empirically attested phonetic universals (descriptive universals), analytic universals, and finally empirical findings that relate descriptive phonetic universals to analytic phonetic universals.

2. Defining phonetic universals

The discussion of linguistic universals can be approached from various perspectives, including syntax and generative grammar, comparative language typology, and phonetic theory. In the Chomskian tradition, two main types of universals have been posited: substantive universals and formal universals. Substantive universals pertain to the fundamental elements of language, such as distinctive features or morphemes. For instance, the presence of the element *vowel* in all spoken languages or *gesture* in all languages exemplifies a substantive universal. On the other hand, formal universals concern the rules governing the combination of these basic elements across languages. Identifying such universals traditionally reflects a commitment to the cognitive representation of language.

An important consideration in crosslinguistic analysis is that the building blocks of language may not be consistent across languages. Nevertheless, establishing a set of generalizable units for comparison, i.e., a meta-language, offers significant advantages. 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, be the most descriptively useful units, capture historical language change processes, etc.? Despite these discussions, having a meta-language is 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 transcriptions, or semantic functions related to prosody. These representations serve as standardized units that can be compared across languages, allowing for the extraction of acoustic or articulatory phonetic measurements. Regardless of the theoretical framework, these abstractions are valuable tools for comparing languages.

Variability in the use of phonetic symbols for describing one or many languages can range from overanalysis to underanalysis in the use of the symbols (Anderson et al., 2023). Some linguists may employ IPA symbols primarily to represent minimal pair contrasts at the phonemic level, potentially abstracting over the range of phonetic realizations (overanalysis). This approach can lead to the loss of phonetic contrasts, which could hinder phonetic measurement and subsequent crosslinguistic comparison. Conversely, other linguists may use IPA symbols to faithfully represent phonetic details, often employing diacritics to account for minor variations (underanalysis). While this approach preserves finer phonetic distinctions, it can complicate comparisons between different linguistic descriptions. Standardizing the use of phonetic symbols is crucial for ensuring consistency in cross-language typological comparisons.

Once we establish a set of comparative dimensions, we can examine how they are distributed across languages. An absolute universal would be a feature present in all languages—this is commonly next to

impossible to establish, as we will never have access to all languages, past and present. Alternatively, universals can be understood statistically, indicating whether a property is more prevalent across languages than expected by chance. These statistical generalizations have also been termed crosslinguistic tendencies (Comrie, 1989; Evans & Levinson, 2010; Bickel, 2015). Examples of this approach in phonetics and phonology include the World Atlas of Linguistic Structures, which categorizes languages based on features like the number of tones or the presence of voiced and voiceless stop contrasts in consonant inventories (Dryer & Haspelmath, 2013; Haspelmath, 2021). Another type of universal discussed in linguistic typology is the implicational universal, which states that if a language has X, then it also has Y. While less common in phonetics due to its more continuous nature, implicational universals are still relevant, especially in phonological studies (Gordon, 2016).

Distinguishing between types of universals can be accomplished through the division of descriptive and analytic universals (Hyman, 2008). Descriptive universals are empirical observations of highly consistent crosslinguistic phonetic patterns, while analytic universals involve the principles or constraints that explain these patterns and are theory dependent. While descriptive universals highlight shared phonetic structures across languages, analytic universals offer explanations for why such surface phonetic variations occur. This article adopts this terminology to provide an overview of the literature on phonetic universals.

3. Methodological considerations

3.1.Sampling and biases

Determining universality in an absolute sense would necessitate data from all world languages, an impossible task. In a distributional sense, universality implies a prevalence of the phenomenon that is greater than chance across languages. However, convenience samples can introduce biases that could skew our understanding of how widespread a phenomenon is. Controlling for biases like genealogical and areal bias in the sample is an important step in concluding universality, and increasing sample size can help with this, provided appropriate statistical correction is implemented (see Miestamo et al., 2016 and Naranjo & Becker, 2022 for further discussion of potential biases in the language sample). To obtain an unbiased representation of languages, a few methods have been proposed. The first is stratification, where the sample contains approximately equal and large numbers of language samples that are representative of their historical and geographic relationships. The second is to use more nuanced statistical methods that can control for non-independence between observations such as hierarchical or mixed-effects regression models.

3.2.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

voice onset time (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 offers 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 amounts of crosslinguistic data. Although online searches for "crosslinguistic phonetics" and "laboratory" yield many relevant studies, they typically involve only two to ten 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 is pre-collected for unrelated purposes, but 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. Nevertheless, researchers are limited by the availability of existing data and processing tools in this approach.

Corpus analysis has 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 automatic 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 (DoReCO: Paschen et al., 2020; VoxAngeles for the UCLA Phonetics Lab Archive: Chodroff et al., 2024; VoxCommunis for Common Voice: Ahn & Chodroff, 2022; VoxClamantis for Wilderness: Salesky et al., 2020). 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 voice onset time 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 ten languages (Hutin & Allassonnière-Tang, 2022), vowel f0 in 16 languages (Ting, Sonderegger, Clayards, & McAuliffe, 2024), and articulation rate in consonants and vowels across eight typologically diverse languages (Lo & Soskuthy, 2023).

Identifying a stratified and representative sample of languages can be further enhanced through this 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, macro-areas, and grammatical features of each language.

4. Descriptive universals

A descriptive universal denotes a consistent crosslinguistic phonetic pattern occurring above chance across languages. In phonetics, several such patterns have been identified (see also Keating, 1984 and Maddieson, 1996b for useful catalogues and descriptions). Table 1 summarizes some putative descriptive universals with providing a high-level overview and a relevant, early, but non-exhaustive set of references.[1] These are first presented here as a simple register; following the presentation of analytic universals, many of these will be further discussed with respect to their empirical support and analytic interpretation.

Phenomenon	Summary	References
Intrinsic vowel f0	Low vowels have a longer f0 than high vowels	Keating (1984)
Intrinsic vowel duration	Low vowels have a longer duration than high vowels	Keating (1984)
Extrinsic vowel duration	Vowels are shorter before or after a voiceless consonant than voiced consonant	Keating (1984), Maddieson (1996b), Coretta (2019)
Vowel duration and syllable structure	Vowels in closed syllables (CVC) are shorter than vowels in open syllables (CV)	Maddieson (1996b)
High vowel devoicing	High vowels are more susceptible to devoicing than low vowels	Maddieson (1996b)
Consonant f0	f0 following voiceless consonants is higher than f0 following voiced consonants	Maddieson (1996b)
Stop place of articulation and closure duration	bilabial stops have longer closure duration than velar stops (or more posterior stops)	Maddieson (1996b)
Stop place of articulation and voice onset time	bilabial stops have a shorter VOT than velar stops (or more posterior stops)	Maddieson (1996b)
Word-final devoicing	Voiced stops are less likely in word/utterance-final position (observation: Word-final devoicing is fairly common among languages that allow obstruents in final position)	Keating (1984)
Vowel-to-vowel coarticulation	Coarticulation from one vowel to another is greater in languages with smaller vowel inventories than those with larger vowel inventories	Manuel & Krakow (1984)
Nasal coarticulation	Vowels adjacent to a nasal consonant will also be partially nasalized resulting in a forwards or backwards 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 towards the end of a phrase particularly relative to their duration in phrase-medial position	Maddieson (1996b)
F0 declination and amplitude declination	F0 and amplitude decrease over the course of an utterance	Maddieson (1996b)
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)

Table 1. A non-exhaustive list of putative descriptive phonetic universals that have been previously discussed in the literature.

5. Analytic universals

The following section employs four broad categories to organize the primary themes accounting for phonetic patterns in the sound systems of the world: automatic effects, contrast and dispersion, economy and uniformity, ease, and sound symbolism.

5.1. Automatic effects

With respect to phonetics and phonology, early views on phonetic realization posited that distinctive features had a universal projection into phonetic space (Chomsky & Halle, 1968). These features, like voicing and nasality, were considered universal language building blocks, organized in a matrix of binary values. The phonetic instantiation of these universal features was then assumed to be determined by the speaker's physiology and motor system.

However, extensive evidence now supports that phonetic realization differs substantially across languages (Disner, 1983; Lisker and Abramson, 1964; Gordon et al., 2002; Fuchs & Toda, 2010; Reidy, 2016). Moreover, the phonetic realization of a speech sound like [s] differs not only across languages (Gordon et al., 2002; Heffernan, 2004; Li et al., 2007; Fuchs & Toda, 2010; Reidy, 2016), but also gender (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, challenging the notion of a universal phonetic realization.

Although not all aspects of phonetic realization are automatic and universal across languages, some dimensions might still be linked to a biomechanical factors involved in the instantiation. In Keating (1984) and Maddieson (1996b), the term *phonetic universal* referred to a biomechanical or automatic consequence of producing a speech sound. This differs from the earlier use of *universal*, in which the observation was more common than what would be expected by chance.

Several putative descriptive universals may stem from automatic physical effects of speech production. For instance, pitch and amplitude declination during speech could result from decreased subglottal pressure over time after the initial breath. Similarly, intrinsic f0 may be explained biomechanically via the tongue-pull hypothesis and associated jaw movement: tongue raising and jaw movement might tighten the cricothyroid muscle, raising f0, akin to tightening a string on a guitar for higher pitch (e.g., Chen et al., 2021 for an overview). A common question is whether such effects are under speaker control or merely by-products of articulation. An implicit assumption in the explanation for automatic effects is the notion that different speech sounds should have the same phonetic targets along some dimensions (e.g., f0 should be the same for /i/ and /a/).[2] This is explored further in Section 5.3 on economy and uniformity.

Relevant to this discussion is the distinction between mechanistic and ecological universals (Maddieson, 1996b). Mechanistic universals arise from the automatic biomechanics of speech articulation, while ecological universals align with analytic factors like contrastivity and connectedness between speech

[1]These are qualified as putative as many have not been thoroughly tested with comparable and large amounts of data across a cross-linguistically diverse set of languages. Nevertheless, evidence has been put forth in support of their existence.

[2] Some might instead argue that automatic effects arise from unspecified aspects of a speech sound, as opposed to an explicitly defined phonetic target. Regardless of whether the phonetic dimension is specified or not, the logic of a biomechanical explanation holds only when the assumption was for the observed dimension to have otherwise been the same across the two speech sounds.

sounds. While biomechanics may explain many effects, Maddieson stresses that ecological explanations are essential, as phonetic patterns are unlikely to be fully accounted for by biomechanics alone.

5.2. 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 to subtle factors like intrinsic vowel f0. While the notion of *maximal* contrast dates back to Jakobson (1941) and has had considerable influence in the field, Lindblom (1986) later relaxed this to emphasize *sufficient* contrast within phonetic inventories.

One of the foundational proposals is that vowel categories should be maximally dispersed within the relevant phonetic space. Liljencrants and Lindblom (1972) introduced a quantitative model that optimized the vowel distribution within the phonetic space. By minimizing the inverse distance between vowels, the model maximizes overall distance of a given vowel inventory size in an F1 x F2' mel space, where F2' is defined in terms of both F2 and F3. Coordinates resulting from simulations for different inventory sizes were then labeled with IPA symbols based on their canonical formant values.

This model predicts vowel inventory structures based solely on maximizing dispersion in the F1 x F2' mel space. 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/; however, the model tends to favor /o/ paired with /ɛ/ given its greater phonetic distance from /o/. This issue is revisited in the subsequent section 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, the Theory of Adaptive Dispersion predicts that within a given language, phonetic variation should be greater in a small vowel inventory than 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/, [æ ɐ a ɑ ɒ] for /a/. Several empirical studies have followed up on this prediction, but with inconclusive results (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 cross-linguistically 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, despite /ɛ ɐ ʊ/ being 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 and quantal theories, Schwartz et al. (1997b) 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 intra-formant 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 x 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 struggles to predict the prevalence of schwa and symmetrical vowel systems. Additionally, the model does not fully consider potential articulatory constraints (e.g., ease of articulation) that could influence vowel preferences.

Extending this, Cotterell & Eisner (2017, 2018) introduced a generative model for vowel inventories that not only addresses principles of dispersion and focalization, but also 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 (1995) and subsequent works implemented Dispersion Theory within an Optimality Theory framework. Three overarching constraints were proposed: 1) maximizing contrast distinctiveness, 2) minimizing articulatory effort, and 3) maximizing the number of contrasts (Flemming, 1995, 2004). This theory balances competing constraints governing phonological inventory structure, including perceptual distinctiveness, feature economy (Section 5.3), and articulatory ease (Section 5.4).

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 utilize 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.

Dispersion theory and quantal vowel regions alone have been argued to be insufficient in accounting 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).

5.3. Economy and uniformity

Another set of analytic phonetic universals focuses on principles of economy, uniformity, symmetry, or reuse or 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 utilize 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, kʼ, ts, ɬ, m, r, |]), languages typically opt for more symmetric and featurally economical systems. Similarly, Clements (2003a, b) proposes Feature Economy, stating that "languages maximize the ratio of sounds over features", predicting, for example, that a language with /p t k/ is more likely to have /b d g/ than /ɖ ɟ ɢ/.

Maddieson (1996a)'s Gestural Economy argues similarly, but at a phonetic level, suggesting that languages or individuals reuse physical gestures across segments. Gestures are considered physical and dynamic, and 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., 2007b; 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", i.e., perceptuo-motor targets, rather than abstract phonological features.

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. The study investigated the phonetic realization of the laryngeal feature in aspirated stop consonants across place of articulation. Some speakers maintained a uniform glottal spreading gesture and timing relationship, while others exhibited near-identical voice onset times between /b d 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.

Rather than directly affecting articulatory or acoustic instantiation, Chodroff & Wilson (2017, 2022) proposed a target uniformity constraint that promotes uniformity in the abstract phonetic targets corresponding to a distinctive feature value. 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. Although some variation from perfect reuse of targets may occur, the constraint aims to minimize this relative to other factors like dispersion or articulatory ease.

Faytak (2018) also argues for a critical role of uniformity in shaping the sound system of a language. The claim is also 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. 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.

5.4. Articulatory ease

In addition to dispersion and uniformity, another constraint influencing phonetic realization is articulatory ease (Lindblom and 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 earlier, 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 (1990)'s 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.

5.5. Sound symbolism

Another potential analytic factor that could account for phonetic variation is sound symbolism. One notable example of this is the frequency code hypothesis in the use of f0 (Ohala, 1983a, b, 1984). Across languages and even species, the use of f0 has many common interpretations: "high f0 signifies (broadly) smallness, nonthreatening attitude, desirous of the goodwill of the receiver, etc., and low f0 conveys largeness, threat, self-confidence, and self-sufficiency." Across languages, questions tend to have a rising f0, whereas statements have falling f0. Tone languages also tend to use high tones for small, diminutive, or narrow objects or concepts, and low tones for large objects or concepts. Across cultures and even species, a high f0 is frequently interpreted as more submissive or nonthreatening, whereas a low f0 is interpreted as more confident, dominant, or aggressive. F0 happens to be closely tied to anatomy across species, which may serve as a physical and evolutionary explanation for these consistent cross-language and cross-species associations with high and low f0.

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 explores various empirical phonetic findings concerning 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 dimensions across two or more speech sounds. For example, while fundamental frequency (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 (1984) 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, or IF0, 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, confirming 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, finding a slight, but non-significant positive correlation. Moreover, this effect has even been found in babbling among English- and French-acquiring infants, suggesting 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. (2023) examined intrinsic f0 and consonant f0 across 16 languages from 9 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 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 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%. 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 and Ladd, 2016), Shona (Gonzales, 2009), Taiwanese (Zee, 1980), various English dialects (Jacewicz and Fox, 2015), and Yoruba (Hombert, 1977). The effect, however, can be modulated by various factors: 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 non-prominent syllables (Ladd and Silverman, 1984; Shadle, 1985; Steele, 1986) and lower pitch ranges (Ladd and Silverman, 1984; Whalen and Levitt, 1995).

6.1.2. Intrinsic vowel duration

Intrinsic vowel duration refers to the observation that low vowels typically having longer durations than high vowels, and tense vowels longer durations than lax vowels (House & Fairbanks, 1953; Peterson & Lehiste, 1960; Lindblom, 1967; Keating, 1984). This effect has been argued to reflect physical factors such as the jaw displacement duration of low vowels relative to high vowels. 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).

Intrinsic vowel duration 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; Gandour, 1984). In a study of American English, Catalan, and Japanese, Solé and 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 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, or CF0, 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, such as voiceless aspirated or unaspirated stops, or 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 was 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. 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, despite both involving 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 historically distinguished by VOT, but now involves both VOT and f0 contrasts. Specifically, aspirated and lenis stops no longer differ in VOT, but 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. (2023) observed a consistent direction in the consonant f0 effect in all 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 and 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 and 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 sharing the same laryngeal status, voice onset time (VOT) shows an inverse relationship with place of articulation: stops with more posterior places tend to have longer absolute VOTs (Fischer-Jorgensen, 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 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 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 include 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 find that only a fixed duration of vocal fold opening adequately explains 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 the 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 was 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 Flege (1989), comparing English (large) to Spanish (small), Jongman et al. (1989), comparing German and English (large) to Greek (small), and 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 and Espinosa (2009) found that smaller vowel systems were no less dispersed than larger ones, and 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 non-peripheral vowels with a larger F2 range. In addition, the phonetic realization of specific vowel categories was often found to be more variable across languages, with the exact realization depending 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 & Allassonièrre-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) utilized 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 high similarity of a given phonetic dimensions across otherwise contrastive speech sounds. In addition to the automatic effects discussed earlier, several studies have also 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 perceptuo-motor 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 constraint, suggests an economy of targets rather than dispersion. The authors argue 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 mid-frequency peaks of [s] and [z] across 18 languages from 6 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, finding that terminal falls were predominant in statements (35 out of 37 languages), terminal rises in polar questions (37 out of 41 languages), and terminal falls 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, with the only exceptions 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 potential universal influences like sound symbolism (Ladd, 1981).

Several empirical studies have also examined similarities and differences in the rhythmic profile of languages. Traditionally, languages have been classified as either syllable-timed or stress-timed in terms of their rhythm, suggesting a universal dichotomy (Pike, 1946; Abercrombie, 1965, 1967; see Grabe & Low, 2003 for an overview). 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 (2003) found 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), Mandarin, Cantonese, and Thai (Dellwo et al., 2014), among others.

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.

Acknowledgements

This work was supported by Grant PR00P1_208460 from the Swiss National Science Foundation.

References

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. *Journal of Phonetics*, *17*(1–2), 47–54.

Ahn, E., & Chodroff, E. (2022). VoxCommunis: A Corpus for Cross-linguistic Phonetic Analysis. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 5286–5294). European Language Resources Association. https://aclanthology.org/2022.lrec-1.566

Anderson, C., Tresoldi, T., Greenhill, S. J., Forkel, R., Gray, R., & List, J.-M. (2023). Variation in phoneme inventories: Quantifying the problem and improving comparability. *Journal of Language Evolution*, lzad011. https://doi.org/10.1093/jole/lzad011

Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., & Weber, G. (2019). Common Voice: A massively-multilingual speech corpus. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 4218–4222.

Arvaniti, A. (2009). Rhythm, Timing and the Timing of Rhythm. *Phonetica*, *66*(1–2), 46–63. https://doi.org/10.1159/000208930

Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S., & Kostadinova, T. (2003). Do rhythm measures tell us anything about language type? *Proceedings of the 15th International Congress of Phonetic Sciences*, 2693–2696.

Becker-Kristal, R. (2010). *Acoustic typology of vowel inventories and Dispersion Theory: Insights from a large cross-linguistic corpus*. UCLA.

Bell-Berti, F. (1975). Control of pharyngeal cavity size for English voiced and voiceless stops. *The Journal of the Acoustical Society of America*, *57*(2), 456–461.

Bertinetto, P. M. (1989). Reflections on the dichotomy 'stress' vs.'syllable-timing.' *Revue de Phonétique Appliquée*, *91*(93), 99–130.

Bickel, B. (2015). Distributional Typology: Statistical Inquiries into the Dynamics of Linguistic Diversity. In *The Oxford Handbook of Linguistic Analysis*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199677078.013.0046

Black, A. W. (2019). CMU Wilderness Multilingual Speech Dataset. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5971–5975. https://doi.org/10.1109/ICASSP.2019.8683536

Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, *25*(2), 217–270. https://doi.org/10.1017/S0952675708001474

Bolinger, D. (1978). Yes—No questions are not alternative questions. In *Questions* (pp. 87–105). Springer.

Bradlow, A. R. (1995). *A comparative acoustic study of English and Spanish vowels*.

Bundgaard, M. (1980). An acoustic investigation of intrinsic vowel duration in Danish. *Annual Report of the Institute of Phonetics University of Copenhagen*, *14*, 99–119.

Chen, Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics*, *39*(4), 612–625. https://doi.org/10.1016/j.wocn.2011.04.001

Chistovich, L. A., & Lublinskaya, V. V. (1979). The "center of gravity" effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research*, *1*, 185–195. https://doi.org/10.1016/0378-5955(79)90012-1

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229. https://doi.org/10.1006/jpho.1999.0094

Chodroff, E., Golden, A., & Wilson, C. (2019). Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. *The Journal of the Acoustical Society of America*, *145*(1), EL109–EL115. https://doi.org/10.1121/1.5088035

Chodroff, E., Pažon, B., Baker, A., & Moran, S. (2024). Phonetic Segmentation of the UCLA Phonetics Lab Archive. *Proceedings of the Fourteenth Language Resources and Evaluation Conference.*

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, *61*, 30–47. https://doi.org/10.1016/j.wocn.2017.01.001

Chodroff, E., & Wilson, C. (2022). Uniformity in phonetic realization: Evidence from sibilant place of articulation in American English. *Language*.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.

Clements, G. N. (2003a). Feature economy as a phonological universal. *Proceedings of the 15th International Congress of Phonetic Sciences*, 371–374.

Clements, G. N. (2003b). Feature economy in sound systems. *Phonology*, *20*(3), 287–333.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, *66*, 185–216. https://doi.org/10.1016/j.wocn.2017.09.009

Cohen Priva, U., Strand, E., Yang, S., Mizgerd, W., Creighton, A., Bai, J., Mathew, R., Shao, A., Schuster, J., & Wiepert, D. (2021). The cross-linguistic phonological frequencies (XPF) corpus. *Providence: Brown University*.

Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology*. University of Chicago press.

Conneau, A., Ma, M., Khanuja, S., Zhang, Y., Axelrod, V., Dalmia, S., Riesa, J., Rivera, C., & Bapna, A. (2023). Fleurs: Few-shot learning evaluation of universal representations of speech. *2022 IEEE Spoken Language Technology Workshop (SLT)*, 798–805.

Connell, B. (2002). Tone languages and the universality of intrinsic F 0: Evidence from Africa. *Journal of Phonetics*, *30*(1), 101–129. https://doi.org/10.1006/jpho.2001.0156

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

Cotterell, R., & Eisner, J. (2017). Probabilistic Typology: Deep Generative Models of Vowel Inventories. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1182–1192. https://doi.org/10.18653/v1/P17-1109

Cotterell, R., & Eisner, J. (2018). A Deep Generative Model of Vowel Formant Typology. In M. Walker, H. Ji, & A. Stent (Eds.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 37–46). Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1004

Cruttenden, A. (1981). Falls and rises: Meanings and universals. *Journal of Linguistics*, *17*(1), 77–91. https://doi.org/10.1017/S0022226700006782

Dellwo, V., Mok, P., & Jenny, M. (2014). Rhythmic variability between some Asian languages: Results from an automatic analysis of temporal characteristics. *Proceedings of Interspeech 2014*, 1708–1711.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*(2), 121–144.

Disner, S. F. (1983). Vowel quality: The relation between universal and language-specific factors [PhD Thesis]. In *UCLA Working Papers in Phonetics* (Vol. 58). PhD Dissertation. UCLA.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77–95. https://doi.org/10.1016/j.wocn.2014.12.005

Docherty, G. (1992). *The timing of voicing in British English obstruents*. Walter de Gruyter.

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology. https://wals.info/

Elert, C.-C. (1964). *Phonologic studies of quantity in Swedish: Based on material from Stockholm speakers*. University of Stockholm.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Faytak, M. (2022). *Place uniformity and drift in the Suzhounese fricative and apical vowels*. *8*(s5), 569–581. https://doi.org/10.1515/lingvan-2021-0071

Faytak, M. D. (2018). *Articulatory uniformity through articulatory reuse: Insights from an ultrasound study of Sūzhōu Chinese* [PhD Thesis]. University of California, Berkeley.

Fischer-Jørgensen, E. (1954). Acoustic analysis of stop consonants. *Le Maître Phonétique*, *32*, 42–59.

Flege, J. E. (1989). Differences in inventory size affect the location but not the precision of tongue positioning in vowel production. *Language and Speech*, *32*(2), 123–147.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (1st ed., pp. 232–276). Cambridge University Press. https://doi.org/10.1017/CBO9780511486401.008

Flemming, E. S. (1995). *Auditory Representations in Phonology* [PhD Thesis, UCLA]. https://linguistics.ucla.edu/general/dissertations/Flemming.1995.pdf

Flemming, E. S. (2001). *Auditory Representations in Phonology*.

Fougeron, C. (1998). *Variations articulatoires en début de constituants prosodiques de différents niveaux en français* [PhD Thesis]. Université Paris 3-Sorbonne Nouvelle.

Francis, A. L., Ciocca, V., Wong, V. K. M., & Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *The Journal of the Acoustical Society of America*, *120*(5), 2884–2895. https://doi.org/10.1121/1.2346131

Fruehwald, J. (2013). *The phonological influence on phonetic change*. University of Pennsylvania.

Fruehwald, J. (2017). The role of phonology in phonetic change. *Annual Review of Linguistics*, *3*, 25–42. https://doi.org/10.1146/annurev-linguistics-011516-034101

Fuchs, S., & Toda, M. (2010). Do differences in male versus female /s/ reflect biological or sociophonetic factors? In S. Fuchs, M. Toda, & M. \.Zygis (Eds.), *Turbulent Sounds. An Interdisciplinary Guide* (pp. 281–302). Walter de Gruyter. https://doi.org/10.1515/9783110226584.281

Gandour, J. (1984). Vowel duration in Thai. *Crossroads: An Interdisciplinary Journal of Southeast Asian Studies*, *2*(1), 59–64.

Gao, J., & Arai, T. (2019). Plosive (de-)voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery. *Journal of Phonetics*, *77*, 100932. https://doi.org/10.1016/j.wocn.2019.100932

Geluykens, R. (1988). On the myth of rising intonation in polar questions. *Journal of Pragmatics*, *12*(4), 467–485. https://doi.org/10.1016/0378-2166(88)90006-9

Gonzales, A. (2009). Intrinsic F0 in Shona vowels: A descriptive study. *Selected Proceedings of the 39th Annual Conference on African Linguistics, Ed. Akinloye Ojo and Lioba Moshi*, 145–155.

Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, *32*(2), 141–174. https://doi.org/10.1017/S0025100302001020

Gordon, M. K. (2016). *Phonological Typology*. Oxford University Press.

Gordon, M., & Roettger, T. (2017). Acoustic Correlates of Word Stress: A Cross-Linguistic Survey. *Linguistics Vanguard*, *3*(1). https://doi.org/10.1515/lingvan-2017-0007

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, *7*(515–546), 1–16.

Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals: Age of acquisition effects on the mutual influence of the first and second languages. *Phonetica*, *60*(2), 98–128.

Guy, G. R., & 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. *Linguistic Typology*, *26*(3), 605–670. https://doi.org/10.1515/lingty-2021-0002

Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2024). *Glottolog 5.0* [dataset]. Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.10804357

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in Englisha). *J. Acoust. Soc. Am.*, *125*(1).

Haspelmath, M. (2021). General linguistics must be based on universals (or non-conventional aspects of language). *Theoretical Linguistics*, *47*(1–2), 1–31. https://doi.org/10.1515/tl-2021-2002

Hauser, I. (2022). Speech sounds in larger inventories are not (necessarily) less variable. *The Journal of the Acoustical Society of America*, *152*(5), 2664–2674. https://doi.org/10.1121/10.0014912

Heffernan, K. (2004). Evidence from HNR that /s/ is a social marker of gender. *Toronto Working Papers in Linguistics*, *23*, 71–84.

Hombert, J.-M. (1977). Consonant types, vowel height and tone in Yoruba. *Studies in African Linguistics*, *8*(2), 173.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic Explanations for the Development of Tones. *Language*, *55*(1), 37. https://doi.org/10.2307/412518

Hoole, P., & Honda, K. (2011). Automaticity vs. Feature-enhancement in the control of segmental F0. In *Where do phonological features come from* (pp. 131–171). John Benjamins BV Amsterdam/Philadelphia.

House, A. S., & Fairbanks, G. (1953). The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels. *The Journal of the Acoustical Society of America*, *25*(1), 105–113. https://doi.org/10.1121/1.1906982

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. https://doi.org/10.3390/languages7030234

Hyman, L. M. (2008). Universals in phonology. *The Linguistic Review*, *25*(1–2). https://doi.org/10.1515/TLIR.2008.003

Jacewicz, E., & Fox, R. A. (2015). Intrinsic fundamental frequency of vowels is moderated by regional dialect. *The Journal of the Acoustical Society of America*, *138*(4), EL405–EL410. https://doi.org/10.1121/1.4934178

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Suhrkamp Frankfurt aM.

Jongman, A., Fourakis, M., & Sereno, J. A. (1989). The acoustic vowel space of Modern Greek and German. *Language and Speech*, *32*(3), 221–248.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, *45*, 76–90. https://doi.org/10.1016/j.wocn.2014.03.005

Keating, P. (1984). Universal phonetics and the organization of grammars. *UCLA Working Papers in Phonetics*, *59*, 35–49.

Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics*, *18*(3), 321–334.

Keating, P. A. (2003). Phonetic and other influences on voicing contrasts. In M. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 20–23). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_0375.pdf

Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. (2004). Domain-initial articulatory strengthening in four languages. *Phonetic Interpretation: Papers in Laboratory Phonology VI*, 143–161.

Kingston, J. (2011). Tonogenesis. In *The Blackwell Companion to Phonology* (pp. 1–30). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444335262.wbctp0097

Kingston, J., & Diehl, R. L. (1994). Phonetic Knowledge. *Language*, *70*(3), 419–454. https://doi.org/10.1353/lan.1994.0023

Kirby, J., Kleber, F., Siddins, J., & Harrington, J. (2020). Effects of prosodic prominence on obstruent-intrinsic F0 and VOT in German. *Speech Prosody 2020*, 210–214. https://doi.org/10.21437/SpeechProsody.2020-43

Kirby, J., & Ladd, D. R. (2018). Effects of obstruent voicing on vowel F0: Implications for laryngeal realism. *Yearbook of the Poznan Linguistic Meeting*, *4*(1), 213–235. https://doi.org/10.2478/yplm-2018-0009

Kirby, J. P., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel *F* 0: Evidence from "true voicing" languages. *The Journal of the Acoustical Society of America*, *140*(4), 2400–2411. https://doi.org/10.1121/1.4962445

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347.

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, *16*(2), 153–169.

Kohler, K. J. (1982). F_0 in the Production of Lenis and Fortis Plosives. *Phonetica*, *39*(4–5), 199–218. https://doi.org/10.1159/000261663

Ladd, D. R. (1981). On Intonational Universals. In T. Myers, J. Laver, & J. Anderson (Eds.), *The Cognitive Representation of Speech* (Vol. 7, pp. 389–397). North-Holland. https://doi.org/10.1016/S0166-4115(08)60214-9

Ladd, D. R., & Schmid, S. (2018). Obstruent voicing effects on F0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops. *Journal of Phonetics*, *71*, 229–248. https://doi.org/10.1016/j.wocn.2018.09.003

Ladd, R., & Silverman, K. E. A. (2009). Vowel Intrinsic Pitch in Connected Speech. *Phonetica*, *41*(1), 31–40. https://doi.org/10.1159/000261708

Ladefoged, P., Blankenship, B., Schuh, R. G., Jones, P., Gfroerer, N., Griffiths, E., Harrington, L., Hipp, C., Jones, P., Kaneko, M., Moore-Cantwell, C., Oh, G., Pfister, K., Vaughan, K., Videc, R., Weismuller, S., Weiss, S., White, J., Conlon, S., … Toribio, R. (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. *Proc. Interspeech 2019*, 3970–3974. https://doi.org/10.21437/Interspeech.2019-2808

Lee, J. L., Ashby, L. F. E., Garza, M. E., Lee-Sikka, Y., Miller, S., Wong, A., McCarthy, A. D., & Gorman, K. (2020). Massively multilingual pronunciation modeling with WikiPron. *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC)*, 4223–4228.

Lehiste, I., & Peterson, G. E. (1961). Some Basic Considerations in the Analysis of Intonation. *The Journal of the Acoustical Society of America*, *33*(4), 419–425. https://doi.org/10.1121/1.1908681

Li, F., Edwards, J., & Beckman, M. (2007). Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. *Proceedings of the Sixteenth International Congress of Phonetic Sciences*.

Liljencrants, J., & Lindblom, B. (1972). Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language*, *48*(4), 839. https://doi.org/10.2307/411991

Lindblom, B. (1967). Vowel duration and a model of lip mandible coordination. *Speech Transmission Laboratory Quarterly Progress Status Report*, *4*, 1–29.

Lindblom, B. (1983). Economy of speech gestures. In *The Production of Speech* (pp. 217–245). Springer.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. Jaeger (Eds.), *Experimental Phonology* (pp. 13–44). Academic Press.

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. M. Hyman & C. Li (Eds.), *Language, Speech, and Mind* (pp. 62–78). Routledge.

Lisker, L., & Abramson, A. S. (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. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS)*, *11*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.3871&rep=rep1&type=pdf

Lo, R. Y.-H., & Sóskuthy, M. (2023). Articulation rate in consonants and vowels: Results and methodological challenges from a cross-linguistic corpus study. *Proceedings of the 20th International Congress of Phonetic Sciences*, 3206–3210.

Löfqvist, A. (1975). Intrinsic and extrinsic Fo variations in Swedish tonal accents. *Phonetica*, *31*(3–4), 228–247.

Luo, Q. (2018). *Consonantal effects on f0 in tonal languages*. Michigan State University.

Maddieson, I. (1984). The effects on F0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis. *Journal of Phonetics*, *12*(1), 9–15. https://doi.org/10.1016/S0095-4470(19)30845-9

Maddieson, I. (1996a). Gestural economy. *UCLA Working Papers in Phonetics*, 1–6.

Maddieson, I. (1996b). Phonetic universals. *UCLA Working Papers in Phonetics*, 160–178.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *The Journal of the Acoustical Society of America*, *88*(3), 1286–1298.

Manuel, S. Y., & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research*, *77*(78), 69–78.

Martinet, A. (1957). *Économie Des Changements Phonétiques: Traité de Phonologie Diachronique* (A. F. Berne, Ed.; Vol. 10). Bibliotheca Romanica.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *INTERSPEECH*, *2017*, 498–502.

Ménard, L., Schwartz, J.-L., & Aubin, J. (2008). Invariance and variability in the production of the height feature in French vowels. *Speech Communication*, *50*(1), 14–28. https://doi.org/10.1016/j.specom.2007.06.004

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? *Proceedings of International Congress of Phonetic Sciences (ICPhS)*, 723–726.

Miestamo, M., Bakker, D., & Arppe, A. (2016). Sampling for variety. *Linguistic Typology*, *20*(2), 233–296. https://doi.org/10.1515/lingty-2016-0006

Mortensen, D. R., Dalmia, S., & Littell, P. (2018). Epitran: Precision G2P for many languages. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Nartey, J. N. A. (1982). *On fricative phones and phonemes: Measuring the phonetic differences within and between languages*. University of California, Los Angeles.

Ohala, J. J. (1979). The contribution of acoustic phonetics to phonology. In *Frontiers of Speech Communication Research* (pp. 355–363).

Ohala, J. J. (1983a). Cross-language use of pitch: An ethological view. *Phonetica*, *40*(1), 1–18.

Ohala, J. J. (1983b). The origin of sound patterns in vocal tract constraints. In *The production of speech* (pp. 189–216). Springer.

Ohala, J. J. (1984). An Ethological Perspective on Common Cross-Language Utilization of F_0 of Voice. *Phonetica*, *41*(1), 1–16. https://doi.org/10.1159/000261706

Oushiro, L. (2019). Linguistic Uniformity in the Speech of Brazilian Internal Migrants in a Dialect Contact Situation. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 686–690). Canberra, Australia: Australasian Speech Science and Technology Association Inc. http://www.assta.org/proceedings/ICPhS2019/papers/ICPhS_735.pdf

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). *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2657–2666. https://aclanthology.org/2020.lrec-1.324

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, *32*(6), 693–703.

Pike, K. L. (1945). *The Intonation of American English*. University of Michigan Press.

Pinget, A.-F., & Quené, H. (2023). Effects of obstruent voicing on vowel fundamental frequency in Dutch. *The Journal of the Acoustical Society of America*, *154*(4), 2124–2136. https://doi.org/10.1121/10.0021070

Pratap, V., Xu, Q., Sriram, A., Synnaeve, G., & Collobert, R. (2020). MLS: A Large-Scale Multilingual Dataset for Speech Research. *Proc. Interspeech 2020*, 2757–2761. https://doi.org/10.21437/Interspeech.2020-2826

Pricop, B., & Chodroff, E. (2024). Consonant f0 effects: A case study on Catalan. *Proceedings of Speech Prosody 2024*.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292.

Recasens, D., & Espinosa, A. (2009). Dispersion and variability in Catalan five and six peripheral vowel systems. *Speech Communication*, *51*(3), 240–258. https://doi.org/10.1016/j.specom.2008.09.002

Reidy, P. F. (2016). Spectral dynamics of sibilant fricatives are contrastive and language specific. *The Journal of the Acoustical Society of America*, *140*(4), 2518–2529. https://doi.org/10.1121/1.4964510

Salesky, E., Chodroff, E., Pimentel, T., Wiesner, M., Cotterell, R., Black, A. W., & Eisner, J. (2020). A Corpus for Large-Scale Phonetic Typology. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4526–4546. https://doi.org/10.18653/v1/2020.acl-main.415

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997a). Major trends in vowel system inventories. *Journal of Phonetics*, *25*(3), 233–253. https://doi.org/10.1006/jpho.1997.0044

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997b). The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics*, *25*(3), 255–286. https://doi.org/10.1006/jpho.1997.0043

Schwartz, J.-L., & Ménard, L. (2019). *Structured idiosyncracies in vowel systems*.

Shadle, C. H. (1985). Intrinsic fundamental frequency of vowels in sentence context. *The Journal of the Acoustical Society of America*, *78*(5), 1562–1567. https://doi.org/10.1121/1.392792

Skirgård, H., Haynie, H. J., Blasi, D. E., Hammarström, H., Collins, J., Latarche, J. J., Lesage, J., Weber, T., Witzlack-Makarevich, A., Passmore, S., & others. (2023). Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances*, *9*(16), eadg6175.

Solé, M.-J., & Ohala, J. J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology 10* (pp. 607–656). DE GRUYTER MOUTON. https://doi.org/10.1515/9783110224917.5.607

Steele, S. A. (1986). Nuclear accent F0 peak location: Effects of rate, vowel, and number of following syllables. *The Journal of the Acoustical Society of America*, *80*(S1), S51–S51. https://doi.org/10.1121/1.2023842

Stevens, K. N. (1989). On the Quantal Nature of Speech. *Journal of Phonetics*, *17*, 3–45.

Stuart-Smith, J., Timmins, C., & Wrench, A. (2003). Sex and gender differences in Glaswegian /s/. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1851–1854.

Ting, C., Clayards, M., Sonderegger, M., & McAuliffe, M. (2023). *The cross-linguistic distribution of vowel and consonant intrinsic F0 effects* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/64nhs

Toivonen, I., Blumenfeld, L., Gormley, A., Hoiting, L., Logan, J., Ramlakhan, N., & Stone, A. (n.d.). *Vowel Height and Duration*.

Ultan, R. (1969). *Some General Characteristics of Interrogative Systems. Working Papers on Language Universals, No. 1.*

Van Hoof, S., & Verhoeven, J. (2011). Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *Journal of Phonetics*, *39*(2), 168–177. https://doi.org/10.1016/j.wocn.2011.02.007

Vaux, B., & Samuels, B. (2015). Explaining vowel systems: Dispersion theory vs natural selection. *The Linguistic Review*, *32*(3). https://doi.org/10.1515/tlr-2014-0028

Watt, D. J. L. (2000). Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated? *Language Variation and Change*, *12*(1), 69–101. https://doi.org/10.1017/S0954394500121040

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *The Journal of the Acoustical Society of America*, *73*(4), 1322–1336.

Westbury, J. R., & Keating, P. A. (1980). Central representation of vowel duration. *The Journal of the Acoustical Society of America*, *67*(S1), S37–S37.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23*(3), 349–366. https://doi.org/10.1016/S0095-4470(95)80165-0

Whalen, D. H., Levitt, A. G., & Goldstein, L. M. (2007). VOT in the babbling of French-and English-learning infants. *Journal of Phonetics*, *35*(3), 341–352.

Whalen, D., Levitt, A. G., Hsiao, P.-L., & Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9-, and 12-month-old French-and English-learning infants. *The Journal of the Acoustical Society of America*, *97*(4), 2533–2539.

Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association*, *33*(2), 165–181. Cambridge Core. https://doi.org/10.1017/S0025100303001270

Zee, E. (1980). Tone and vowel quality. *Journal of Phonetics*, *8*(3), 247–258. https://doi.org/10.1016/S0095-4470(19)31474-3

Zhu, J., Yang, C., Samir, F., & Islam, J. (2024). The taste of IPA: Towards open-vocabulary keyword spotting and forced alignment in any language. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics*.

Zhu, J., Zhang, C., & Jurgens, D. (2022). ByT5 model for massively multilingual grapheme-to-phoneme conversion. *Proc. Interspeech 2022*, 446–450. https://doi.org/10.21437/Interspeech.2022-538