Consonant f0 effects: A case study on Catalan

Bogdan Pricop[1] and Eleanor Chodroff[2]

[1]University of Zurich, Department of Comparative Language Science
[2]University of Zurich, Department of Computational Linguistics
bogdan.pricop@uzh.ch, eleanor.chodroff@uzh.ch

Abstract

Consonant f0 (CF0), the phenomenon in which the vowel onset f0 tends to be higher following a voiceless consonant than following a voiced consonant, has been commonly observed across the world's languages. The present study examined this effect in vowels following stop consonants of Catalan, using one of the largest language-specific datasets available in the Mozilla Common Voice Corpus, which after filtering, contained over 150 hours of validated speech data from over 1000 speakers. The study investigated the magnitude of CF0 in Catalan in initial and late portions of the vowel, the linguistic and social factors that modulate this effect (e.g., stop place of articulation, following segment voice, stress, utterance position, phonetic voicing, gender, and dialect), and the degree of speaker variability within the language. While the effect is small and demonstrates considerable inter-speaker variability, our results nevertheless confirm that the CF0 effect is robust in the phonetic realization of Catalan voiced and voiceless stops in the initial and late portions of the vowel. Moreover, the effect of phonetic voicing goes in the opposing direction to CF0, suggesting that the phonetic targets corresponding to f0 are controlled during phonetic realization.

Index Terms: consonant f0, microprosody, Catalan, corpus phonetics

1. Introduction

Fundamental frequency (f0) assumes myriad roles within a language: it can covary simultaneously with prosodic, lexical, and even segmental properties. With respect to segmental covariation, f0 has been shown to vary intrinsically with segmental properties [1, 2], but also extrinsically, where neighboring segments influence the instantiation of f0 [3, 4, 5, 6]. A well-known microprosodic effect on f0 is consonant f0 (CF0), in which the initial f0 of a vowel following a voiceless obstruent is higher than that of a vowel following a voiced obstruent [7, 8, 5]. Systematic investigation of the CF0 effect has been conducted across a wide range of languages [9, 10, 11, 12, 13], and has been found to occur not only between truly voiced and voiceless unaspirated obstruents [14], but also voiceless unaspirated and aspirated stops [15], as well as voiced and voiceless sonorants [16]. Nevertheless, the degree to which such f0 perturbations are automatic consequences of the articulation or are under the control of the speaker has been debated considerably in the literature [17, 18, 19, 20].

This paper seeks to investigate the magnitude, trajectory, and consistency of CF0 across speakers, linguistic factors beyond stop voice (e.g., stop place of articulation, following segment voice, stress, utterance position, and phonetic voicing), and social factors (gender and dialect) in a large-scale study of Catalan. The data is sourced from Mozilla Common Voice [21] and reflects one of the largest available language datasets at the time of writing with over 3000 recorded hours. The depth of this dataset enabled us to refine the data to recordings that had been externally validated for content and to speakers who had provided demographic information such as dialect and gender. The dataset further provided good coverage of vowels following voiced and voiceless stop consonants in a wide variety of contexts in Catalan. This investigation provides insight on the roles of automaticity, phonetic enhancement, and uniformity in the phonetic realization of stop–vowel sequences.

2. Background

2.1. Automatic, controlled and hybrid view on CF0

Differences between phonetic representations have been proposed to be motivated by automatic or controlled factors, so-called *phonetic knowledge* [19]. Automatic phonetic instantiation would be determined by biomechanical constraints on the production of a given segment, where acoustic differences between segments would be predictable based on the physical properties of the vocal tract and the mechanical constraints of the articulators. In contrast, the controlled view states that speakers exhibit an amount of control over their articulators and can choose from a range of alternative realizations limited by constraints. Speakers could use this control to minimize articulatory effort [22], maximize the perceptual distinctiveness of sounds [23], or to maximize phonetic uniformity of a shared feature across segments [24].

CF0 has been argued to arise from automatic effects of obstruent production [17]. During the production of voiceless sounds, the vocal folds are stiffened to prevent phonation, which increases the frequency of the vibration, i.e. f0 [18]. Alternatively, f0 could be lowered in the proximity of voiced stops and not raised in the proximity of voiceless stops, due to vertical vocal fold tension (i.e., lowering of the larynx) [5].

In contrast to an automatic account, speakers could actively control f0 to enhance the contrast between voiced and voiceless stops for improved listener perception. Kingston and Diehl [19] proposed that speakers specifically lower the f0 values of sounds following voiced obstruents. Similarly, Hanson [13] posited that the heightened f0 of vowels following voiceless stops is meant to "strengthen the perceptual saliency of a voicing contrast" between the preceding consonant and the vowel. Evidence for this comes from perceptual experiments showing that synthesized stops were more likely to be perceived as voiceless when followed by a vowel with a high f0 [25, 26].

Nevertheless, an automatic explanation of the effect remains important for understanding the consistency of the effect's direction. Electromyographic data of the cricothyroid (CT) muscle from German speakers indicate significant activation in the voiceless consonant segment with smaller activation found in the vocalic part (confirming claims in [17]). The authors conclude that consonant articulation—and not the active enhancement of auditory properties—had the biggest influence on voicing contrast effects in f0. However, some speakers also had high CT activation throughout the vowel, suggesting some agency over the activation of the CT [27].

Though previous accounts have argued for a hybrid approach [27], the assumption is that automaticity would occur as a late-stage process as a byproduct of articulation, after the phonetic targets have been specified via controlled factors such as enhancement. The automatic explanation, however, assumes uniform phonetic targets across segments. It may indeed be that speakers intentionally maximize uniformity in the phonetic targets during the phonetic realization of a feature value [28, 24]. This constraint would ensure that the mapping between a feature value and its phonetic target is the same (uniform) for all segments with that feature value (i.e., keep the phonetic target for f0 the same for all [+voice] segments). Uniformity could serve as a counterweight to enhancement and potentially counteract late-stage automatic effects.

2.2. Crosslinguistic studies

Many studies converged on the validity of CF0 [16, 13, 14, 29] but until recently, true large-scale investigation of this phenomenon has been limited. In recent years however, increased access to large speech corpora and greater computational power have motivated an ever-growing number of cross-linguistic phonetic studies, including the study of CF0. For example, Ting et al. [30] investigated vowel intrinsic f0 (VF0) and CF0 effects in 16 languages using a large corpus of read speech. As in previous studies, the study found consistent CF0 effects for each language; however, the magnitude of the CF0 effect varied across languages. The authors proposed that the variation in CF0 might be related to the role that it plays in sound change processes like tonogenesis. Given the consistency in the direction, but variation in magnitude, the authors posited a combination of automatic and controlled factors, i.e. a hybrid approach [27], to account for the CF0 effects.

2.3. Catalan

Catalan is a Romance language spoken in Spain in the regions of Catalonia, the Balearic Islands, Valencia, Aragon and Murcia, as well as in Andorra, Italy (in Alghero) and France (Roussillon) by over 9 million speakers. Its phonemic inventory comprises 23 distinct consonants (among which the stop consonants /p, t, k, b, d, g/), 7 phonemic vowels (/i, ɛ, e, a, o, ɔ, u/) and one allophonic vowel (/ə/) [31]. Catalan is a “true voice” language, meaning that the laryngeal contrast between obstruents is generally realized through the presence or absence of vocal fold vibration [32]. This contrast can, however, vary by context: voiceless stops are more likely to lenite with increased voicing in intervocalic position, and voiced obstruents are completely devoiced in word-final position [33]. Catalan also shows vowel reduction (in Eastern varieties) in unstressed syllables with /e/, /ɛ/ and /a/ reducing to /ə/ and with /o/ and /ɔ/ reducing to u [31].

2.4. Goals of the current study

For the current analysis, we focused our investigation on Catalan, which happens to have one of the largest language datasets in the Mozilla Common Voice project [21]. This study examines the magnitude and duration of the CF0 effect across more than 1000 speakers from four regional Catalan varieties. This large-scale corpus analysis allows us to investigate individual differences, along with the interactions of the voice effect with several linguistic and social factors including phonetic voicing, place of articulation, the following segment voice, sentence position, stress, gender, and region. Moreover, we consider the relative roles of automaticity and control in accounting for CF0, as well as the potential competing influences of phonetic enhancement of contrasting features (e.g., [±voice] of preceding stops), and phonetic uniformity of shared features (e.g., [+voice] of vowels).

If CF0 is an automatic, biomechanical consequence of the laryngeal setting in articulation, we should expect a transient effect that does not persist into the vowel, and a negative relationship between onset f0 and the degree of phonetic voicing in the phonologically voiced or voiceless stop. The articulations that either prevent or enable vocal fold vibration are the same ones argued to respectively raise or lower f0 (e.g., vocal fold tension or lowering the larynx). If the effect is under speaker control, it is possible there are simultaneous, competing pressures to maximize contrast [19] or to maximize uniformity of the feature underlying f0 [24]. If speakers actively enhance the [voice] contrast, an f0 difference should be present at least at the phonological level (phonologically voiced vs voiceless), but not necessarily the phonetic level (the amount of phonetic voicing); it might also extend beyond the initial stages of the vowel. If speakers maximize uniformity of the phonetic target for f0 across vowels, then the effect should be overall small, which could result in minor f0 contrasts in both directions.

3. Methods

Catalan recordings were acquired from the Mozilla Common Voice corpus, an online open access corpus with recording data sets for over 100 languages [21]. The recordings consist of read speech produced by users with their available equipment. Recordings can also be validated by other users to ensure that the speakers produced the intended sentence. For our analysis, we employed the Catalan Common Voice Corpus 15 containing over 2500 hours of validated speech from over 35,000 speakers.

Following the procedure implemented for the VoxCommunis Corpus [34], the validated subset of the Catalan Common Voice Corpus was force-aligned using the Montreal Forced Aligner [37]. Forced alignment using the MFA requires an acoustic model and pronunciation lexicon. The acoustic model was developed using a 300-hour subset of the corpus from speakers with over 50 recordings. For the creation of the pronunciation lexicon, the Catalan lexemes were converted to a canonical phonetic representation using the XPF grapheme-to-phoneme (G2P) tool [35]. Stress marking was obtained from a separate Catalan pronunciation lexicon [36].

F0 measurements were then extracted from vowels preceded by a stop consonant /p, t, k, b, d, g/). Specifically, 11 f0 measurements were taken at 10 equally spaced intervals across the vowel, starting at the vowel onset using the Praat To Pitch function with a pitch floor of 75 Hz, a ceiling of 600 Hz, and a time step of 0.01 seconds. The Voice report function was used for the estimation of phonetic voicing within a given interval. The pitch range was set to 75–600 Hz, the maximum period was 1.3, the amplitude factor 1.6, the silence threshold 0.03, and the voicing threshold 0.45.

A stringent filtering pipeline was then implemented to control for potential social variation, as well as measurement error. As we aimed to investigate the initial and long-term influences of preceding stop voice on f0, our filtering mostly targeted the f0 measurement at the first 10% point of the vowel, with additional filtering also applied to f0s on the remaining tokens at the 90% point of the vowel.

Given known vowel reduction in the low and mid vowels in many regional dialects [38], along with the fact that f0 covaries with vowel height [39, 1, 40], the analysis was limited to the high vowels /i/ and /u/, vowels with a duration of at least 50 ms, vowels that were not followed by another vowel or glide, words which had stress labeling, and utterances with at least four words. In addition, only speakers who explicitly identified as male or female were retained, given the influence of sex and gender on f0 [41, 42]. The analysis was also limited to speakers who had indicated coming from one of four well-represented regional areas: the Balearic, Central, Northwestern, or Valencian dialect. (Note that each speaker self-identified as speaking Catalan with one of the aforementioned dialects.) The Balearic and Central dialects are considered Eastern Catalan varieties; the Northwestern and Valencian dialects are considered Western Catalan varieties [31]. Tokens were only retained when the sentence had been validated by at least two external listeners and had not been rejected by any listeners for reading errors. Only speakers who had at least 10 vowel tokens per stop voice category were then retained.

As measurement extraction error was likely, we additionally removed tokens beyond 1.5 standard deviations of the speaker mean and then retained speakers with only normal f0 distributions, as determined by a Kolmogorov-Smirnov test with an alpha of 0.05. We again retained only those speakers with at least 10 vowel tokens per stop voice category. This procedure was applied to both the initial and late f0 analyses. Following this pipeline, a total of 129,344 tokens from 1464 speakers were available for the initial f0 analysis, taken from the 10% point of the vowel. The late f0 analysis, taken from the 90% point of the vowel, included a total of 86,989 tokens from 1087 of the same speakers in the initial f0 analysis.

Table 1: *The number of female and male speakers per dialect, and the number of initial f0 tokens following phonologically voiced and voiceless stops per dialect.*

Dialect	Female	Male	Voiceless	Voiced
Balearic	31	46	3192	3455
Central	578	597	54553	51807
Northwestern	48	69	4346	4128
Valencian	22	73	4148	3715

For the analysis, f0 was measured in hertz and semitones, where semitones were derived with a reference of 50 Hz. Two linear mixed-effects regressions were implemented predicting initial f0 in semitones in the first analysis and late f0 in semitones in the second. The fixed effects were stop voice (voiceless, voiced), stop place of articulation (coronal, dorsal, labial), following segment voice (voiceless, voiced), stress (stressed, unstressed), utterance position (start time of target word / utterance duration; %), the amount of phonetic voicing in the targeted stop (%), region (Balearic, Central, Northwestern, Valencian), and gender (female, male). Note that phonologically voiceless stops had a particularly wide range of phonetic voicing (mean = 42%, median = 33%) [33]. Phonologically voiced stops were dominantly voiced, but still had a full range of observed phonetic voicing (mean = 82%, median = 100%). We additionally included the two-way interactions of stop voice with place of articulation, following segment voice, stress, utterance position, and phonetic voicing, as well as the full interactions between stop voice, region, and gender. The random effects were a by-participant intercept and slopes for stop voice, following segment voice, and for the initial analysis only, their interaction. Further random-effect structure did not converge. Categorical predictors were sum-coded with the final-listed level treated as the held-out level.

4. Results

As shown in Figure 1, the speaker-specific f0 means following a voiceless stop were numerically higher than those following a voiced stop for the initial and late positions for both genders. Though numerically present, the effect was small. As shown in Figure 2, the effect varied considerably by speaker: approximately 26.5% of speakers had a numerically lower initial f0 following a voiceless stop than a voiced stop, and 32.7% of speakers in the late f0 condition. Although speakers differed in the magnitude and direction of the difference, these deviations were minimal, and the relationship between f0s following voiced and voiceless stops was predictable: the voice-specific f0 means were strongly correlated across speakers in initial and late positions (initial: $r(1461) = 0.98$, late: $r(1084) = 0.98$; each $p < 0.001$).

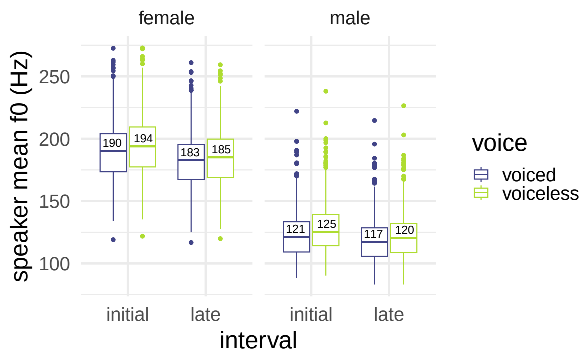

Figure 1: *The speaker-specific mean f0s (Hz) following phonologically voiced and voiceless stops in initial and late positions for female and male speakers. The labeled number reflects the median speaker-specific mean f0 (Hz) for the category.*

The mixed-effects analyses of initial and late f0 revealed a similar pattern of significance, where significance was assessed as a *t*-value greater than the absolute value of 2. We report here the results from the initial and late f0 semitone models. Both analyses revealed a higher f0 following voiceless than voiced stops; however, this effect was considerably larger in the initial than late f0 position (initial: $\beta = 0.643$; late: $\beta = 0.461$). F0 also varied by place of articulation in the initial and late positions: on average, f0 was lower following coronal stops and higher following dorsal stops (initial: $\beta_{coronal} = -0.094$, $\beta_{dorsal} = 0.162$; late: $\beta_{coronal} = -0.145$, $\beta_{dorsal} = 0.295$). In initial position, f0 was lower when the following segment was voiceless than voiced (initial: $\beta = -0.027$); however, in late position, the more expected pattern was observed: f0 was higher when the following segment was voiceless than voiced (late: $\beta = 0.024$). In addition, f0 was higher in stressed than unstressed syllables (initial: $\beta = 0.643$; late: $\beta = 0.461$), and f0 declination was significant, as revealed by utterance position (initial: $\beta = -0.033$; late: $\beta = -0.030$). In contrast to expectations, f0 significantly increased as the percent of phonetic voicing increased in the stop (initial: $\beta = 0.008$; late: $\beta = 0.006$).

Significant interactions were also observed with stop voice, indicating that the size of the voice effect depended on several phonological and social factors. In both initial and late positions, the voice effect was significantly smaller following coronal stops (initial: $\beta = -0.190$; late: $\beta = -0.053$), and in initial position only, significantly larger following dorsal stops (initial: $\beta = 0.132$; late: $\beta = 0.009$, *n.s.*). In late position only, f0 was higher before a voiceless segment and lower before a voiced segment (initial: $\beta = 0.002$, *n.s.*; late: $\beta = 0.027$). In addition, the voice effect significantly decreased with utterance position (initial: $\beta = -0.004$; late: $\beta = -0.003$). In initial position, voice did not interact significantly with the amount of phonetic voicing, indicating only main effects of each (initial: $\beta = 0.0002$, *n.s.*); however, in late position, the voice effect was smaller with increased phonetic voicing (late: $\beta = -0.001$).

With respect to social variables, female speakers had significantly higher f0s than male speakers (initial: $\beta = 3.86$; late: $\beta = 3.66$), and the CF0 effect was significantly smaller in female than male speakers (initial: $\beta = -0.082$; late: $\beta = -0.70$). Region and its interactions with stop voice and gender did not reach significance in the initial or late f0 analyses.

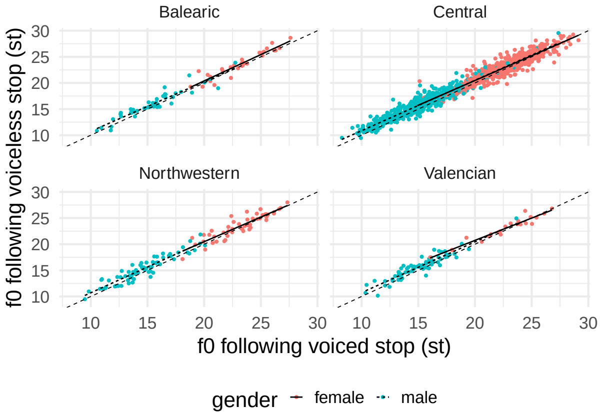

Figure 2: *The paired speaker-specific initial mean f0s following phonologically voiced and voiceless stops in semitones (st), along with the best-fit line of regression. The diagonal dashed line represents the line of identity (y = x).*

5. Discussion and conclusion

The findings of this study align with previous observations in the literature regarding CF0 effects. The results demonstrate that the f0 values of Catalan vowels are significantly influenced by the voicing properties of the preceding obstruents. The CF0 effect was significant not only in the initial, but also late position in the vowel. Moreover, the size of the effect was significantly modulated by linguistic factors of stop place, following segment voice, stress, and utterance position. The lower f0 following coronal stops and higher f0 following dorsal stops have also been numerically observed in previous studies [7, 15]. The smaller effect size with later sentence positions is in line with observations in the literature for CF0 [13] but also for intrinsic vowel f0 [40, 43]. Finally, strong prosodic environments have also been previously reported to support an increase in f0 [44]. Though region was not significant, the effect of CF0 was significantly larger among male speakers relative to female speakers, even with a logarithmic warping to approximate a perceptual scale [cf., 45].

With respect to CF0, a major question in the literature has been whether the effect is a biomechanical consequence of the laryngeal setting for the preceding obstruent, or if in fact speakers control this difference. Though the direction of the effect is consistent with an automatic origin, two important findings speak against a fully automatic effect. First, the CF0 effect was significant not only at the 10% interval of the vowel, but also at the 90% interval. Second, though phonologically voiceless stops indeed had a higher onset f0 than phonologically voiced stops, the more phonetic voicing a stop had—regardless of its phonological category, the *higher* the onset f0 was.

The evidence instead suggests a fair amount of speaker control. In particular, the categorical nature of the CF0 effect is consistent with perceptual enhancement, particularly in combination with the opposing effects of phonetic voicing. If the stop was phonologically voiceless, but realized with 100% voicing, the onset f0 was higher than average; whereas, if the stop was phonologically voiced, but realized with 0% voicing, the onset f0 was lower. Moreover, given that CF0 varies significantly by place of articulation at initial and late positions, speakers could be enhancing place of articulation contrasts in addition to the voicing contrast: dorsal stops tend to be followed by a higher f0 (mean initial f0 = 163 Hz) relative to labial and coronal stop (mean initial f0 for each: 155 Hz).

While the enhancement via CF0 is significant, it is nevertheless small in magnitude, and not always consistent across speakers. This could suggest that some speakers counteract a late-stage automatic effect in phonetic realization to ensure uniformity in the phonetic output of the [+voice] feature of the vowel, regardless of its neighboring segments (i.e., keep f0 the same across vowels) [28, 24].

The large amount of data for Catalan allowed for a precise investigation of gender, dialect, and contextual influences on CF0 in a true voice language. Though the current analysis benefited from this data, several limitations remained. We were unable to code for exact prosodic position in the utterance, instead using a rough approximation with percentage start time within the utterance. Further future directions include investigating variation by vowel height and frontness, the full vowel f0 trajectory, and the extent to which the effect is a raising or lowering one with a comparison to preceding sonorants [16]. The investigation could also be extended to additional languages to assess the degree of language specificity in CF0 [see also 30].

By making use of a large, freely accessible spoken corpus, we were able to investigate the magnitude, trajectory, and consistency of the CF0 effect in over 1000 Catalan speakers. Overall, the results of our study conform with known observations of CF0 cross-linguistically, and indicate a small, but significant difference in the f0 following phonologically voiced and voiceless stops, not only at the vowel onset, but also well into the following vowel.

6. Acknowledgements

The authors would like to thank Annie Baker and the anonymous reviewers for helpful feedback. This project was supported by SNF Grant PR00P1_208460 to EC.

7. References

[1] D. H. Whalen and A. G. Levitt, "The universality of intrinsic f0 of vowels," *Journal of Phonetics*, vol. 23, no. 3, pp. 349-366, 1995.

[2] C. H. Shadle, "Intrinsic fundamental frequency of vowels in sentence context," *The Journal of the Acoustical Society of America*, vol. 78, pp. 1562–1567, 1985.
[3] K. J. Kohler, "F0 in the production of fortis and lenis plosives," *Phonetica*, vol. 39, pp. 199–218, 1982.
[4] K. J. Kohler, "F 0 in the perception of lenis and fortis plosives," *The Journal of the Acoustical Society of America*, vol. 78, no. 1, pp. 21-32, 1985.
[5] J. M. Hombert, J. J. Ohala, and W. G. Ewan, "Phonetic explanations for the development of tones," *Language*, pp. 37-58, 1979.
[6] R. N. Ohde, "Fundamental frequency as an acoustic correlate of stop consonant voicing," *The Journal of the Acoustical Society of America*, vol. 75, no. 1, pp. 224-230, 1984.
[7] A. S. House and G. Fairbanks, "The influence of consonant environment upon the secondary acoustical characteristics of vowels," *The Journal of the Acoustical Society of America*, vol. 25, no. 1, pp. 105-113, 1953.
[8] I. Lehiste and G. E. Peterson, "Some basic considerations in the analysis of intonation," *The Journal of the Acoustical Society of America*, vol. 33, no. 4, pp. 419-425, 1961.
[9] V. Jeel, "An investigation of the fundamental frequency of vowels after various consonants, in particular stop consonants," *Annual Report of the Institute of Phonetics University of Copenhagen*, vol. 9, pp. 191-211, 1975.
[10] S. A. Jun, "Influence of microprosody on macroprosody: a case of phrase initial strengthening," *UCLA Working Papers in Phonetics*, pp. 97-116, 1996.
[11] O. Dmitrieva, F. Llanos, A. A. Shultz, and A. L. Francis, "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*, vol. 49, pp. 77–95, 2015.
[12] A. F. Pinget and H. Quené, "Effects of obstruent voicing on vowel fundamental frequency in Dutch," *The Journal of the Acoustical Society of America*, vol. 154, no. 4, pp. 2124-2136, 2023.
[13] H. M. Hanson, "Effects of obstruent consonants on fundamental frequency at vowel onset in English," *The Journal of the Acoustical Society of* America, vol. 125, no. 1, pp. 425-441, 2009.
[14] J. P. Kirby and D. R. Ladd, "Effects of obstruent voicing on vowel F 0: Evidence from "true voicing" languages," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 2400– 2411, 2016.
[15] Y. Lai, C. Huff, J. Sereno, and A. Jongman, "The raising effect of aspirated prevocalic consonants on F0 in Taiwanese," in *Proceedings of the 2nd International Conference on East Asian Linguistics*, vol. 2, 2009.
[16] I. Maddieson, "The effects on f0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis," *Journal of Phonetics*, vol. 12, no. 1, pp. 9-15, 1984.
[17] M. Halle and K. Stevens, "A note on laryngeal features," *Quarterly Progress Report*, Cambridge, MA: Research Laboratory of Electronics, 101, 198–212, 1971.
[18] A. Löfqvist, T. Baer, N. S. McGarr, and R. S. Story, "The cricothyroid muscle in voicing control," *The Journal of the Acoustical Society of America*, vol. 85, no. 3, pp. 1314-1321, 1989.
[19] J. Kingston and R. L. Diehl, "Phonetic knowledge," *Language*, vol. 70, no. 3, pp. 419-454, 1994.
[20] S. J. Keyser and K. N. Stevens, "Enhancement and overlap in the speech chain," *Language*, pp. 33-63, 2006.
[21] R. Ardila, M. Branson, K. Davis, M. Kohler, J. Meyer, M. Henretty, R. Morais, L. Saunders, F. Tyers, and G. Weber, "Common Voice: A massively-multilingual speech corpus," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, pp. 4218–4222, 2020.
[22] B. Lindblom, "Explaining phonetic variation: A sketch of the H&H theory," *Speech Production and Speech Modelling*, Dordrecht: Springer Netherlands, pp. 403–439, 1990.
[23] J. Liljencrants and B. Lindblom, "Numerical simulation of vowel quality systems: The role of perceptual contrast," *Language*, pp. 839-862, 1972.
[24] E. Chodroff and C. Wilson, "Uniformity in phonetic realization: Evidence from sibilant place of articulation in American English," *Language*, vol. 98, no. 2, pp. 250-289, 2022.
[25] M. Haggard, S. Ambler, and M. Callow, "Pitch as a voicing cue," *The Journal of the Acoustical Society of America*, vol. 47, pp. 613–617, 1969.
[26] O. Fujimura, "Remarks on stop consonants: Synthesis experiments and acoustic cues," *Form and substance: phonetic and linguistic papers presented to Eli Fischer-Jørgensen*, Odense: Akademisk Forlag, pp. 221-232, 1971.
[27] P. Hoole and K. Honda, "Automaticity vs. feature-enhancement in the control of segmental f0," *Where do phonological features come from*, pp. 131-171, 2011.
[28] P. Keating, "Phonetic and other influences on voicing contrasts," in *Proceedings of the 15th international congress of phonetic sciences*, pp. 20-23, Aug. 2003.
[29] J. Kirby and D. R. Ladd, "Effects of obstruent voicing on vowel f0: Implications for laryngeal realism," *Yearbook of the Poznan Linguistic Meeting*, vol. 4, no. 1, pp. 213–235, 2018.
[30] C. Ting, M. Clayards, M. Sonderegger, and M. McAuliffe, "The cross-linguistic distribution of vowel and consonant intrinsic f0 effects," 2023.
[31] J. F. Carbonell and J. Llisterri, "Catalan," *Journal of the International Phonetic Association*, vol. 22, no. 1-2, pp. 53-56, 1992.
[32] J. Beckman, M. Jessen, and C. Ringen, "Empirical evidence for laryngeal features: Aspirating vs. true voice languages," *Journal of linguistics*, vol. 49, no. 2, pp. 259-284, 2013
[33] J-I. Hualde and J. Zhang, "Intervocalic lenition, contrastiveness and neutralization in Catalan", *Isogloss. Open Journal of Romance Linguistics*, vol.8, no. 4, pp. 1-20, 2022.
[34] E. Ahn and E. Chodroff, "VoxCommunis: A corpus for cross-linguistic phonetic analysis," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 5286-5294, Jun. 2022.
[35] U. Cohen Priva, E. Strand, S. Yang, W. Mizgerd, A. Creighton, J. Bai, and D. Wiepert, "The cross-linguistic phonological frequencies (XPF) corpus," *Providence: Brown University*, 2021, https://cohenpr-xpf.github.io/XPF/Convert-to-IPA.html (accessed Dec. 2023).
[36] TALN Natural Language Processing Research Group – UPF, https://github.com/TalnUPF/phonetic_lexica/tree/master/ca (accessed Dec. 2023).
[37] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable text-speech alignment using kaldi," *Interspeech*, pp. 498-502, Aug. 2017.
[38] J. Nogués-Graell, "Vowel Reduction in Catalan Varieties. Catalan typologies and property analysis," Master thesis, UiT Norges Arktiske Universitet, 2019.
[39] H. C. Taylor, "The fundamental pitch of English vowels," *Journal of Experimental Psychology*, vol. 16, no. 4, p. 565, 1933.
[40] C. H. Shadle, "Intrinsic fundamental frequency of vowels in sentence context," *The Journal of the Acoustical Society of America*, vol. 78, no. 5, pp. 1562-1567, 1985.
[41] I. R. Titze, "Physiologic and acoustic differences between male and female voices," *The Journal of the Acoustical Society of America*, vol. 85, no. 4, pp. 1699-1707, 1989.
[42] S. P. Whiteside, "Sex-specific fundamental and formant frequency patterns in a cross-sectional study," *The Journal of the Acoustical Society of America*, vol. 110, no. 1, pp. 464-478, 2001.
[43] R. Ladd and K. E. Silverman, "Vowel intrinsic pitch in connected speech," *Phonetica*, vol. 41, no. 1, pp. 31-40, 1984.
[44] J. Kirby, F. Kleber, J. Siddins, and J. Harrington, "Effects of prosodic prominence on obstruent-intrinsic F0 and VOT in German", In *Proceedings of Speech Prosody*, pp. 210-214, May. 2020.
[45] F. Nolan, "Intonational equivalence: an experimental evaluation of pitch scales," in *Proceedings of the 15th international congress of phonetic sciences*, vol. 771, no. 4, Aug. 2003.