# STOP VOICING AND F0 PERTURBATIONS: EVIDENCE FROM FRENCH AND ITALIAN

*James P. Kirby and D. Robert Ladd*

Linguistics and English Language, University of Edinburgh
j.kirby@ed.ac.uk, bob.ladd@ed.ac.uk

## ABSTRACT

We report new experimental evidence on consonant-induced F0 perturbations in two languages with prevoiced stops, French and Italian. A positive correlation between duration of voicing lead and F0 at the onset of post-release voicing is observed, consistent with the predictions of an automatic or biomechanical account of the source of this effect. While the findings do not strictly rule out a role for onset F0 as a controlled enhancement, they support the proposal that, if anything, the enhancement is of [-voice] or [stiff] rather than [+voice].

**Keywords**: voicing, onset F0, enhancement, French, Italian

## 1. INTRODUCTION

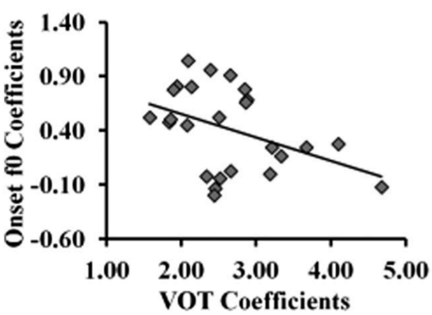
In a wide variety of languages, it has been documented that vowel fundamental frequency (F0) at the onset of periodic voicing tends to be lower after voiced than after voiceless consonants [9, 14, 20]. We will refer to this phenomenon, which has gone by a number of different names, as the *onset voicing effect* (OVE). Onset F0 has been shown to be accessible as a cue to the voicing specification of consonants (usually stops), especially under conditions when other cues are supressed or ambiguous [24, 26], and has also been implicated in the emergence of lexical tone systems [1, 9].

While the existence of the OVE is well established, the ultimate source of the effect has not yet been resolved. Early accounts of the OVE proposed that it was an automatic, biomechanical (e.g., myoelastic or aerodynamic) concomitant of the mechanics of voicing [7]. The downward movement of the larynx, which facilitates voicing by increasing supraglottal volume while also decreasing stiffness of the vocal tract, is naturally accompanied by a shortening and slackening of the vocal folds, which translates to lower fundamental frequency; similarly, raising the larynx stiffens the vocal folds, inhibiting vocal-fold vibration and manifesting as higher F0 on the following vowel [10, 12].

One challenge for purely biomechanical accounts is that the OVE is found both in 'true voicing' languages such as Dutch or French, where phonologically [+voice] stops are phonetically (pre)voiced [22], as well as in languages such as English and German, where the phonetic contrast between stops (at least in initial position) is typically one of aspiration, without accompanying vocal-fold vibration [2]. This and other observations have led to the proposal that the OVE is instead the result of a controlled strategy to enhance the *phonological* contrast between obstruents [16]. On this account, the similarity in cross-linguistic behaviour follows from the assumption that languages like French and English share a feature [voice], irrespective of any other aspects of its phonetic implementation. Speakers may enhance this contrast through actively lowering the pitch of the [+voice] member of the opposition, with the goal of efficiently producing a composite perceptual cue structured from more basic acoustic properties [16, 17].

Recent work by Shultz et al. [25] is consistent with this hypothesis. Data from a production study of 32 native speakers of American English showed a significant inverse correlation between voice onset time (VOT) and onset F0 (Figure 1), suggesting that the OVE was attenuated by speakers who produced stops with longer VOTs. The authors note that, in addition to being consistent with the essentially auditory account of the OVE offered by [16, 17], their finding is also what would be expected if VOT and onset F0 are in a trading relationship, whereby a change in the value of one cue that would otherwise result in a different phonetic percept can be offset by a change in the value of another [24]. It is also what would be expected if the magnitude of enhancement gestures varies relative to a cue's robustness in signalling a phonetic contrast [6, 18].

**Figure 1**: Relation between onset F0 and VOT coefficients in production (from [25]).



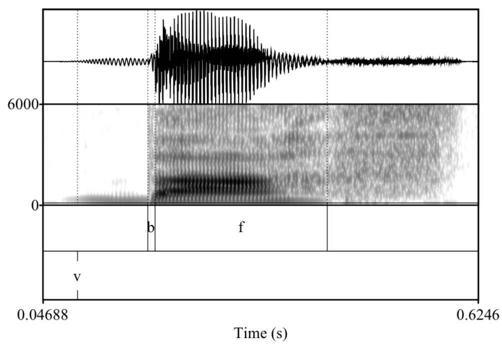
The explicit connection between the OVE and the phonological specification [+voice] drawn by Kingston & Diehl [16] led us to ask if a similar relation might obtain in languages with *pre*voiced stops as well. If the OVE is indeed a controlled enhancement of [+voice], we expect to find a negative correlation similar to that seen by Shultz et al. [25], whereby the shorter the voicing lead, the lower the F0 at vowel onset. To investigate this we conducted a similar experiment with two 'true voicing' languages, French and Italian.

## 2. EXPERIMENT

Six female speakers each of French and Italian participated in this study. Each speaker was recorded reading a list of isolated test words beginning with /p/ and /b/ and a list of the same words preceded by articles or prepositions, which placed the test consonant in intervocalic position. Each randomly-ordered list contained 3 repetitions drawn from a list of 10 minimal or near-minimal triplets involving common words: two bilabial target items (e.g. *balla* ~ *palla* in Italian, or *bar* ~ *part* in French) and one rhyming distractor (e.g. *stalla, lard*; see Appendix).

Following segmentation (see Figure 2), VOT was measured as the difference between the onset of periodic voicing and the release of stop closure, and F0 was extracted at 7 equidistant timepoints over the vowel (the interval marked 'f' in Figure 2).

**Figure 2**: Segmentation illustrated with initial syllable of *basta* 'enough', initial position. First tier shows (b)urst and duration of (f)ollowing vowel; second tier shows onset of (v)oicing.



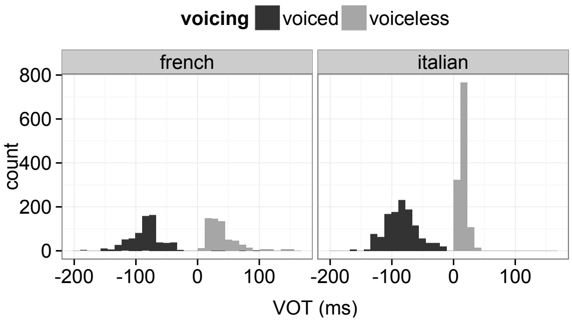

## 3. RESULTS

All statistical analyses presented below describe linear mixed-effect models fit in R 3.1.2 [23] using the *lme4* [3] and *lsmeans* [21] packages.

### 3.1. VOT

Figure 3 shows the distribution of VOT by language and voicing. As expected, in both languages, phonologically [+voice] stops are characterised by extensive prevoicing (mean of -85ms with standard deviation of 26ms in both languages). For [-voice] stops, French stops were both more aspirated (mean VOT 40 ms) and this aspiration was more variable (s.d. 26 ms) compared to Italian voiceless stops (mean VOT 15 ms, s.d. 6 ms). This finding is consistent with recent work indicating that French voiceless plosives may have an active specification for glottal spreading [13].

**Figure 3**: Distribution of VOT in French (left) and Italian (right) by voicing category.




### 3.2. F0 contour

Figure 4 shows the time course of F0 over the vowel F0 (measured at 7 equidistant timepoints) by phonation type and context in both languages, averaged over items and speakers. Although French participants tended to produce the test words with a rising-falling intonation contour and Italian participants predominantly used a steadily falling contour, the OVE is clearly present at post-release voicing onset in both languages and in both contexts.

Differences in F0 at voicing onset and vowel midpoint were assessed statistically using two pairs of linear mixed-effect analyses (at voicing onset and vowel midpoint) with random intercepts for subjects and items (using the Satterthwaite approximations to degrees of freedom for calculating *p*-values). Within-context pairwise comparisons support the interpretation that F0 is higher at voicing onset following [-voice] stops in both languages and in both contexts (all adjusted $p < 0.01$), but that there is no difference in F0 at vowel midpoint (note that error bars in Figure 4 show standard error of the mean rather than standard deviation).

### 3.3. VOT-F0 covariance

Figure 5 shows the covariance of VOT and F0 by language, phonation type and contrast measured at the onset of post-release voicing. In these plots, each point represents the VOT-F0 relationship at the

**Figure 4**: Time course of F0 realisation by language, context (initial/medial) and voicing type. Error bars show standard error of the mean; x-axis shows normalized time.

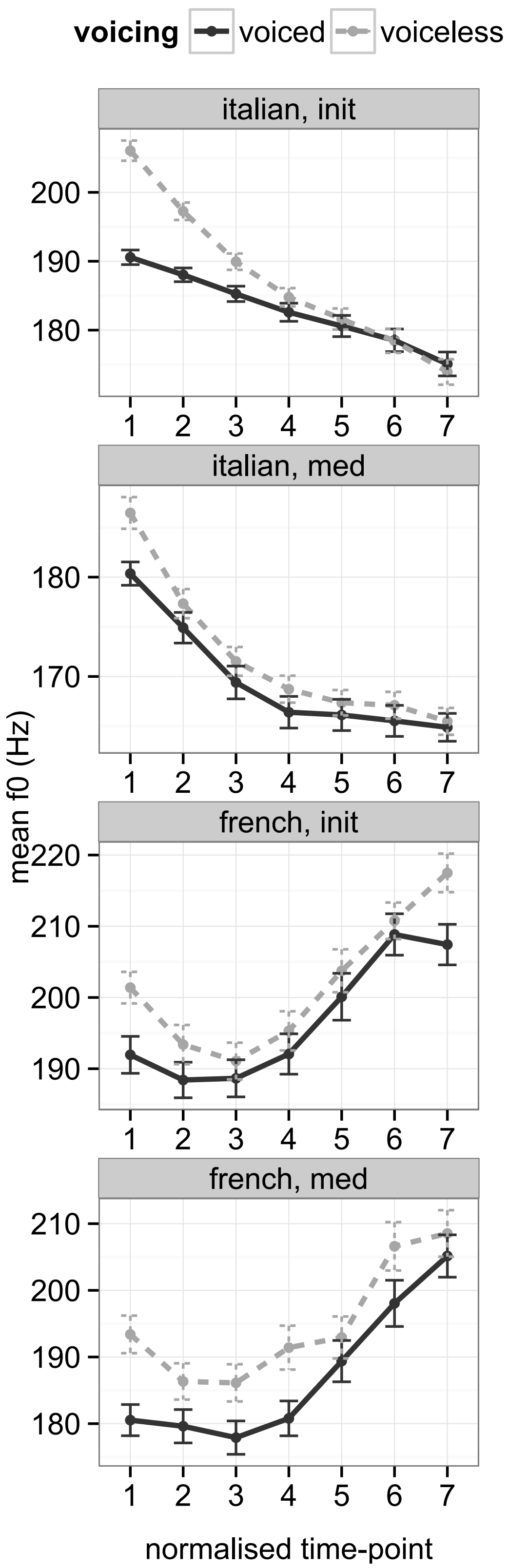
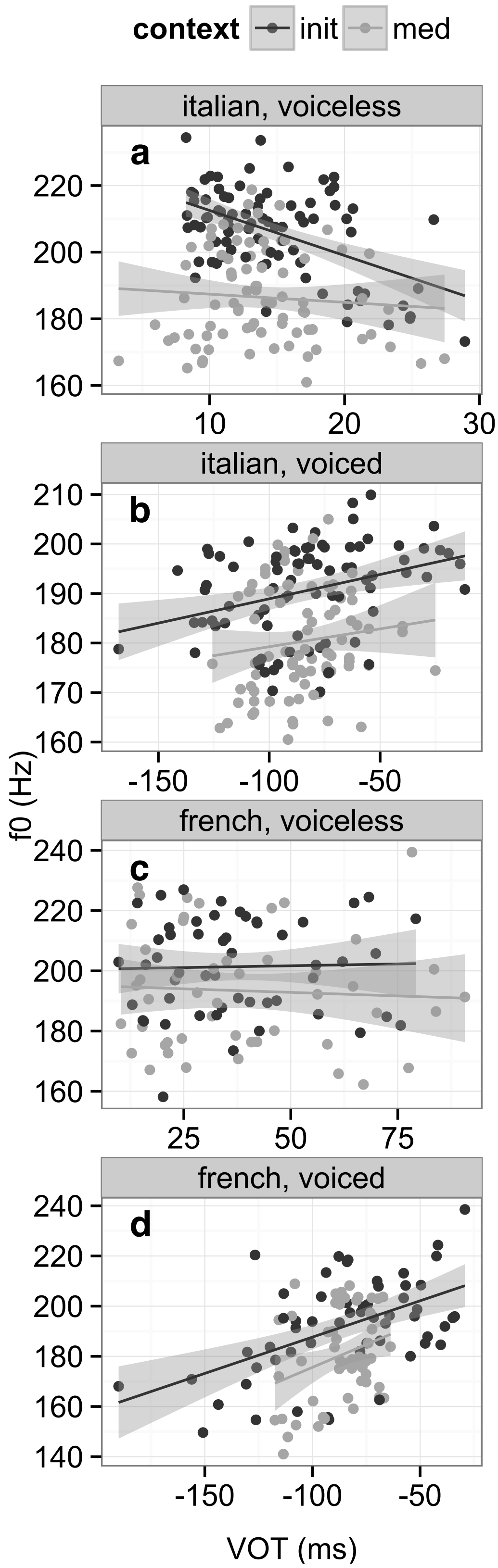



**Figure 5**: VOT-F0 relationship by language, context (initial/medial) and voicing type. Lines indicate linear trends modelled by context; shading indicates 95% confidence interval.



onset of periodic post-release voicing for a single production token. This relationship was also assessed by predicting onset F0 from VOT, positional Context, Voicing type, and the interaction of VOT with both factorial predictors in a pair of mixed-effects models with random intercepts for subjects and items (Tables 1-2). In Italian, VOT-F0 covariance for voiceless stops shows an inverse relationship similar to that found by Shultz et al. [25]: for longer VOTs, onset F0 was lower in both initial and medial contexts (Table 1 and Figure 5, panel a). In French, no covariance between VOT and onset F0 was observed for voiceless stops in either context (Table 2 and Figure 5, panel c). For voiced stops, however, longer voicing lead was accompanied by *lower* onset F0 in both languages and in both contexts (Figure 5, panels b and d).

**Tables 1-2**: Coefficient and standard error estimates, Satterthwaite approximated degrees of freedom and *t* statistics predicting F0 at post-release onset from VOT, voicing category, position in word (context) and their interactions. Both models included random intercepts for subjects and items, with VOT centred in both cases. Reference levels: Voice=voiced; Context=initial.

*Table 1: Italian*

| | Estimate | SE | df | *t* | Pr(>\|t\|) |
|---|---|---|---|---|---|
| (Intercept) | 195.18 | 4.53 | 5.3 | 43.07 | **<0.001** |
| VOT | 0.08 | 0.03 | 311.7 | 2.79 | **<0.01** |
| Voice:voiceless | 30.47 | 8.00 | 317.6 | 3.81 | **<0.001** |
| Context:medial | -15.12 | 0.90 | 308.4 | -16.7 | **<0.001** |
| VOT:voiceless | -0.49 | 0.16 | 316.5 | -3.01 | **<0.01** |
| VOT:medial | -0.08 | 0.02 | 307.5 | -4.71 | **<0.001** |

*Table 2: French*

| | Estimate | SE | df | *t* | Pr(>\|t\|) |
|---|---|---|---|---|---|
| (Intercept) | 197.26 | 7.78 | 4.96 | 25.36 | **<0.001** |
| VOT | 0.09 | 0.04 | 200.1 | 2.48 | **<0.05** |
| Voice:voiceless | -4.66 | 4.46 | 38.34 | -1.05 | 0.30 |
| Context:medial | -9.64 | 1.34 | 199.8 | -7.19 | **<0.001** |
| VOT:voiceless | 0.06 | 0.07 | 74.55 | 0.85 | 0.40 |
| VOT:medial | 0.02 | 0.02 | 201.8 | 0.78 | 0.44 |

## 4. DISCUSSION

Our data from both French and Italian speakers indicate that, rather than an inverse correlation between voicing lead and onset F0, longer prevoicing is accompanied by *lower* F0 at the onset of post-release voicing. This is exactly what would be expected if the OVE is a by-product of larynx lowering as a strategy to facilitate pharyngeal expansion and maintain the transglottal pressure differential necessary to maintain voicing during the stop closure [4, 27]. Although voicing lead and onset F0 may indeed trade perceptually in these languages, our data do not provide evidence that speakers actively seek to trade one against the other in production, cf. [25].

In Italian, we observed an inverse VOT-onset F0 correlation for [-voice] stops, similar to that seen by Shultz et al. [25] in American English. That this effect was not also found in our French data may be a function of differences in production of the global intonation contour. In our Italian data, where speakers produced targets with predominantly falling intonation, longer VOT would translate to more time for the presumably stiffened vocal folds to relax and approximate the (low) intonational F0 target, producing an inverse VOT-F0 correlation. In French, any such effect may have been overridden by the rising-falling intonation produced by our participants. We are conducting a follow-up study with more careful controlled intonational contexts in order to better understand the effects of this difference, if any, on VOT-onset F0 covariance.

While our present findings are broadly supportive of a primarily automatic account of the OVE, they are not necessarily incompatible with an enhancement account. Indeed, while it is likely the case that the OVE is ultimately of biomechanical origin, this in no way rules out the possibility of speakers deliberately exaggerating the effect for enhancement purposes [5, 11, 15, 19]. What our present findings do suggest is that, cross-linguistically, the (phonetic) target of this enhancement is more likely to be the (phonologically) *voiceless* member of the opposition [7, 8, 15]. We are currently seeking additional evidence on the controlled vs. automatic nature of the OVE in French and Italian using the paradigm of Hanson [8] to compare the size and extent of the OVE in different intonational contexts with respect to a sonorant baseline.

## 5. ACKNOWLEDGMENTS

The authors would like to thank Maria Dokovová and Signe Jorgensen for their assistance with preparing the data for acoustic analysis. This project was supported in part by a grant from the University of Edinburgh Moray Endowment.

## 6. APPENDIX: WORDLIST

*French:* bar-part-lard, beau-peau-chaud, beurre-peur-soeur, bière-Pierre-lierre, boule-poule-foule

*Italian*: Bacco-pacco-stacco, balla-palla-stalla, banca-panca-stanca, banda-panda-Standa, basta-pasta-casta

## 7. REFERENCES


[1] Abramson, A. S. 2004. The plausibility of phonetic explanations of tonogenesis. In: Fant, G., Fujisaki, H., Cao, J., Xu, Y. (Eds.), *From traditional phonology to modern speech processing: Festschrift for Professor Wu Zongji's 95th birthday*. Beijing: Foreign Language Teaching and Research Press, 17–29.

[2] Abramson, A. S., Lisker, L. 1968. Voice timing: Cross-language experiments in identification and discrimination. *J. Acoust. Soc. Am.* 44(1), 377.

[3] Bates, D., Maechler, M., Bolker, B. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, http://CRAN.R-project.org/package=lme4.

[4] Bell-Berti, F. 1975. Control of pharyngeal cavity size for English voiced and voiceless stops. *J. Acoust. Soc. Am.* 57(2), 456–461.

[5] Chen, Y. 2011. How does phonology guide phonetics in segment–f0 interaction? *J. Phon.* 39(4), 612–625.

[6] Cox, F., Palethorpe, S. 2014. Phonologisation of vowel duration and nasalised /æ/ in Australian English. *Proceedings of the 15th Australasian Speech Science & Technology Conference*, Christchurch, 33–36.

[7] Halle, M., Stevens, K. N. 1971. A note on laryngeal features. *MIT Q. Prog. Rep.* 101, 198–212.

[8] Hanson, H. M. 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *J. Acoust. Soc. Am.* 125(1), 425.

[9] Hombert, J.-M., Ohala, J. J., Ewan, W. G. 1979. Phonetic explanations for the development of tones. *Language* 55(1), 37–58.

[10] Honda, K., Hirai, H., Kusakawa, N. 1993. Modeling vocal tract organs based on MRI and EMG observations and its implication on brain function. *Res. Inst. Logop. Phoniatr. Annu. Bull.* 27, 37–49.

[11] Hoole, P., Honda, K. 2011. Automaticity vs. feature-enhancement in the control of segmental F0. In: Clements, N., Ridouane, R. (Eds.), *Where do phonological features come from?: Cognitive, physical and developmental bases of distinctive speech categories*. Amsterdam: John Benjamins, 133–171.

[12] Hoole, P., Honda, K., Murano, E., Fuchs, S., Pape, D. 2004. Cricothyroid activity in consonant voicing and vowel intrinsic pitch. *Proceedings of the Conference on Voice Physiology and Biomechanics*, Marseille.

[13] Hoole, P., Bombien, L. 2014. A cross-language study of laryngeal-oral coordination across varying prosodic and syllable-structure conditions. *Proceedings of the 10th International Seminar on Speech Production (ISSP)*, Cologne, 198–201.

[14] House, A. S., Fairbanks, G. 1953. The influence of consonant environment upon the secondary acoustic characteristics of vowels. *J. Acoust. Soc. Am.* 25(1), 105–113.

[15] Keyser, S. J., Stevens, K. N. 2006. Enhancement and overlap in the speech chain. *Language* 82(1), 33–63.

[16] Kingston, J., Diehl, R. L. 1994. Phonetic knowledge. *Language* 70(3), 419–454.

[17] Kingston, J., Diehl, R. L., Kirk, C. J., Castleman, W. A. 2008. On the internal perceptual structure of distinctive features: The [voice] contrast. *J. Phon.* 36(1), 28–54.

[18] Kirby, J. 2013. The role of probabilistic enhancement in phonologization. In: Yu, A. C. L. (Ed.), *Origins of sound patterns: approaches to phonologization*. Oxford: Oxford University Press, 228–246.

[19] Kohler, K. J. 1982. F0 in the production of fortis and lenis plosives. *Phonetica* 39, 199–218.

[20] Lehiste, I., Peterson, G. E. 1961. Some basic considerations in the analysis of intonation. *J. Acoust. Soc. Am.* 33(4), 419–425.

[21] Lenth, R., Hervé, M. 2014. *lsmeans: Least-Squares Means*. R package version 2.13, http://CRAN.R-project.org/package=lsmeans.

[22] Löfqvist, A., Baer, T., McGarr, N. S., Storey, R. S. 1989. The cricothyroid muscle in voicing control. *J. Acoust. Soc. Am.* 85, 1314–1321.

[23] R Core Team. 2014. *R: A language and environment for statistical computing* (version 3.12). Vienna, Austria: R Foundation for Statistical Computing.

[24] Repp, B. H. 1982. Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychol. Bull.* 92(1), 81–110.

[25] Shultz, A. A., Francis, A. L., Llanos, F. 2012. Differential cue weighting in perception and production of consonant voicing. *J. Acoust. Soc. Am.* 132(2), EL95–EL101.

[26] Silverman, K. 1986. F0 segmental cues depend on intonation: the case of the rise after voiced stops. *Phonetica* 43, 76–91.

[27] Westbury, J. R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *J. Acoust. Soc. Am.* 73(4), 1322–1336.