COVARIATION OF ACOUSTIC CUES FOR VOICING IN APHASIA AND APRAXIA[1]

Anna Marczyk[1] & Maria-Josep Solé[2]

[1] Aix-Marseille Université, CNRS, LPL, Aix-en-Provence, France
[2] Universitat Autònoma de Barcelona, Spain

ABSTRACT

We report experimental evidence on covariation between vowel onset *f*0 and voice onset time (VOT) in voiced and voiceless stops produced by Spanish speakers with apraxia of speech and conduction aphasia as compared to a healthy control group. These two disorders are argued to affect different components of the same dorsal stream involved in mapping sounds onto motor-based representations. On the assumption that acoustic cue covariation reflects—albeit indirectly—trade-off relationships in speech production, we explore how these deficits affect compensatory articulation. Our findings show a trade-off relationship between *f*0 and VOT for voiced but not voiceless stops in the control and conduction aphasia groups, and in some apraxic speakers. Interestingly, the apraxic group shows compensatory cue correlation in devoiced stops. We relate observed patterns in cue trade-off to the internal structure of phonetic categories and compensatory articulation.

Keywords: compensation, apraxia, aphasia, voiced stops.

1. INTRODUCTION

It is well established that the voicing contrast is realized by means of multiple acoustic correlates. Recent studies testify to a growing interest in the interplay or 'trade-off' among acoustic cues in signaling the voicing contrast, based on the underlying assumption that such compensatory relationships are guided by phonological specifications (i.e., actively controlled) rather than being mere mechanical effects of articulatory adjustments for voicing [8, 12, 13, 14]. The present paper is in line with these studies in that it investigates the trade-off between vowel onset *f*0 and voice onset time (VOT) in the stop voicing contrast in Spanish. On the assumption that speech movements are programmed to achieve acoustic/auditory goals [18] and that consequently the trade-off between cues observed in the acoustic signal should reflect similar interactions in the articulatory domain (i.e., motor equivalence [19]), we examine and compare cue correlations in the production of stops by healthy and aphasic speakers. In particular, we seek to determine whether speakers with neurologically-based deficits that interfere with processes of word-form encoding, namely speakers with apraxia of speech (AOS) and conduction aphasia (CA), engage in compensatory behavior that surfaces as a trade-off relationship between acoustic cues. In the next paragraphs we lay out the theoretical motivations for our hypotheses.

AOS and CA are two major aphasic syndromes that share a number of characteristics but also differ in crucial aspects of speech production. Both disorders affect the sound shape of words and exhibit so-called 'phonemic paraphasias', that is, errors which surface as segmental substitutions (e.g., *gata* 'cat' identified as /ˈkata/) or transposition of phonemes (the same word identified as /'taɣa/). Both populations are typically aware of their deviant productions and tend to make frequent attempts at correction. However, the symptomatologies of the two clinical pictures differ in that, among other things, the speech of CA speakers tends to be fluent and well-articulated, in contrast to the slow, laborious, and aprosodic speech characteristic of AOS.

From a functional point of view, it is commonly thought that voicing errors (e.g., /g/ > [k]) may arise at different levels in the two clinical groups [5, 6, 16]: at the level of phonetic programming in non-fluent AOS speakers and at the level of phonological encoding in fluent CA speakers. Distinguishing between these two categories has traditionally been based on VOT measurements. In particular, it has been suggested that frequent devoicing errors in AOS are due to difficulties in synchronizing events that belong to two different subsystems, the laryngeal and supralaryngeal, as captured by VOT [2]. Yet, if the VOT values land outside the voiced category, we hypothesize that AOS speakers may emphasize so-called secondary cues to voicing, showing compensatory articulation. Thus, an examination of the subphonemic structure of speech sounds may throw new light on the phonetics vs. phonology dichotomy in aphasic speech errors.

Because this paper revolves around the issue of compensatory articulation, we will frame our hypotheses within an internal monitoring view, inclusive of phonetic corrective mechanisms. While several influential speech production models feature

mechanisms of interaction between motor commands and error detection components, the Hierarchical State Feedback Control model (HSFC) [10] offers a parsimonious account of errors that arise in both AOS and CA within one single architecture (for comparison, the DIVA model [9] would account for apraxic errors while GODIVA [3] would account for CA errors). According to HSFC, AOS and CA affect different components of the same dorsal stream, which is involved in mapping sounds onto motor-based representations. The model posits that CA, which is associated with damage to the temporoparietal junction, disrupts sensorimotor integration and consequently prevents the internal correction mechanisms from generating a feedforward corrective command. Accordingly, we should expect speakers with CA to either exhibit abnormal patterns of compensatory articulation or not exhibit it at all. This prediction is consistent with acoustic data from patients with fluent aphasia, including CA, which exhibit phonetic irregularities termed 'subtle phonetic impairment' [21].

AOS, on the other hand, is associated with a disruption of feedforward control presumably due to left frontal cortical lesions with putative involvement of subcortical, insular areas. Because an intermittent access to mental representations is a well agreed-upon feature of aphasic speech, including AOS, researchers generally assume that on-target productions reflect a correct retrieval of phonological specifications for a given speech sound. Alternatively, we hypothesize that such items are correctly perceived thanks to compensatory articulation, specifically, the enhancement of secondary acoustic cues when the VOT values fail to signal the voicing contrast. If this is the case, we should observe an increase in compensatory activity in correctly identified stops, as captured by negative correlations between VOT and secondary cues, as well as by the steepness of the regression slope relative to that for healthy speakers.

The predictions for errors are somewhat less straightforward. In the case of devoicing errors (/g/ → [k]), there is a clear mismatch between the auditory and motor outputs, because the error involves crossing the phonological—and frequently also lexical—boundary (e.g., *gata* and *cata* are different words in Spanish). There is currently no consensus among researchers as to whether compromising phonological and lexical value should trigger greater compensatory activity [4, 7, 11, 17]. The current study examines the covariation between VOT and vowel onset *f*0 in word-initial Spanish stops produced by CA, AOS, and healthy speakers to ascertain if speakers with speech impairments exhibit compensatory articulation.

2. METHODS

2.1. Subjects, materials and procedure

Six apraxic speakers, four conduction aphasic speakers, and six healthy speakers took part in this study. Three out of six apraxic speakers had mild or residual aphasic deficit. All were native speakers of Spanish or bilingual Spanish-Catalan speakers with Spanish self-reported as the predominant language. The clinical subjects were classified into apraxia and conduction aphasia on the basis of an initial assessment using the MTBA battery [15] adapted to Peninsular Spanish, which confirmed a speech therapist's preliminary assessment. Control speakers matched the participants in terms of age, gender, and socio-cultural and sociolinguistic characteristics.

The subjects were instructed to read and repeat words containing voiced and voiceless stops in stressed and unstressed phrase-initial position (e.g., /'gata/ vs. /ga'naɾ/). The following vowel was always /a/. The number of voiceless/voiced stops analyzed per speaker group was 218/182 (control), 207/171 (AOS), and 122/109 (CA). Acoustic and auditory analyses were conducted. First, words were transcribed in orthography by two investigators working independently and each stop was classified as 'successful voiceless/voiced' or 'unsuccessful'. The category 'unsuccessful' involved in all cases voiced stops identified as voiceless. In the rest of the paper we will refer to intended voiced stops perceived as voiceless as 'devoiced'. Acoustic analyses were then performed for the successful voiced and voiceless stops, and unsuccessful stops for control, AOS and CA speakers. A Praat script was used to extract VOT and vowel onset *f*0 at the first available pulse and no later than 6 pulses into the vowel.[2] Raw onset *f*0 data were normalized using a z-score transformation to allow for comparison across subjects.

2.2. Statistical analyses and interpretation of slopes

To evaluate the relationship between VOT and secondary features, multiple linear mixed-effects models (lme4 package) were built for correctly and incorrectly identified stops for each group of speakers. The first set of analyses tested whether *f*0 varied depending on the voicing category (voiced, 'devoiced', voiceless). Next, we built separate models for each voicing category and group (3 categories × 3 groups) with *f*0 as the response variable and VOT as predictor. Subject and item were random factors. Cue covariation was frequently nonlinear and best fitted with a higher-term equation. In the case of quadratic equation (1), we report the values of the regression coefficient β_2, which informs about the direction and

steepness of the curvature (convexity being associated with positive coefficients and concavity with negative coefficients).

(1) $y = \beta_0 + \beta_1(x) + \beta_2(x^2)$

The covariation between cues to signal voicing contrasts in normal and clinical speech may show multiple patterns. To help the interpretation of regression slopes, we describe three main patterns. A *negative* correlation of the cues suggests a trade-off between the two cues. That is, when one cue is weak or sub-optimal, the other cue is realized at a higher degree of optimality, as seen, for example, when in voiced stops short negative VOTs (sub-optimal for Spanish voiced stops) and a lower onset *f*0 co-occur and longer negative VOTs and a high onset *f*0 co-occur. Negatively correlated VOT and *f*0 would indicate that 'overall informativity' is maintained but the two cues are given different weights [1, 12]. Lack of correlation between two cues indicates that the two cues do not co-vary. A *positive* correlation may indicate that two cues co-vary mechanically due to articulatory dependencies; alternatively, it may suggest redundancy, or a reinforcing of the cue's informativity. Because our focus of study is clinical vs. healthy speech, if the same correlations between two cues are found across groups, these patterns may be a general property of cue covariation, rather than a different weighting of cues when VOT is impaired.

3. RESULTS

The results of the auditory analysis showed that speakers with AOS made significantly more devoicing errors than the other two groups. Specifically, 38.6% (n = 66) of the intended voiced stops produced by the speakers with AOS were identified as voiceless compared with only 15.6% (n = 17) for CA, and 2.19% (n = 4) for the control group. Voiceless stops were correctly identified for all groups. Comparison of the auditory and the acoustic data in Fig. 1 reveals that 'devoiced' stops and on-target voiceless stops had similar +VOT values for the two clinical groups (and the control group), whereas correctly identified voiced stops showed prevoicing.

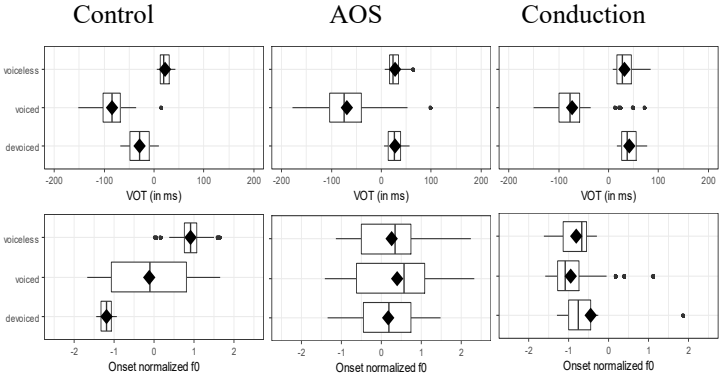
3.1. VOT

The top panels in Fig. 1 show the distribution of VOT for correctly identified voiceless and voiced stops, and for 'devoiced' stops by group. For all three groups, correctly identified voiced stops show extensive prevoicing (M = -85ms (4); M = -76ms (12); M = -81ms (6) for control, AOS, and CA respectively), with a larger variance for the clinical groups than the control group. As expected, all three groups also show voiceless stops with short +VOT values (M = 14 ms (5); M = 31 ms (12); M = 34 ms (8)). Phonetic category (voiceless, voiced, devoiced) predicted variation in VOT for all three groups (control, $F_{(2,98)} = 187$, $p = .000$; AOS, $F_{(2,98)} = 136$, $p = .000$; CA $F_{(2,163)} = 77$, $p = .000$). Post-hoc analyses indicated that VOT for voiced vs. voiceless stops differed significantly for all three groups of speakers ($F_{(1,24)} = 415$, $p = .000$), but no difference was found between voiceless and devoiced stops in any of the groups.

3.2. Onset *f*0

The bottom panels in Fig. 1 show the distribution of onset *f*0 by phonetic category and speaker group. Statistical analysis showed a significant effect of phonetic category on *f*0 variation only in the control group ($F_{(1,109)} = 6.68$, $p = .011$), with voiced stops showing a significantly lower *f*0 at vowel onset (M = -.16 Hz (0.39) than voiceless stops (M = .02 Hz (39), p = .000). Devoiced stops showed lower *f*0 than both voiced and voiceless stops (M = -2.11 Hz (.5)) but due to a very small sample size (only four tokens) they were not included in the model. The *f*0 differences failed to reach significance in either the AOS ($p = .059$) or the CA ($p = .879$) group.

Figure 1. VOT (top) and vowel onset *f*0 (bottom) distribution across phonetic categories and groups.



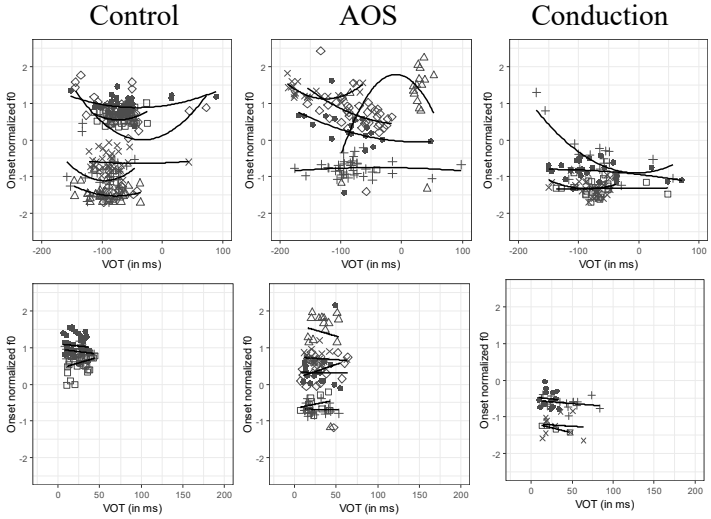
3.3. Cue covariation: Onset *f*0 and VOT

Fig. 2 shows the covariation of *f*0 and VOT by group for on-target voiced (top) and voiceless stops (bottom). In these plots, each point represents the VOT-*f*0 relationship for a single token. The different symbols represent individual speakers. Because the group data obscures the individual patterns, regression lines are drawn for each speaker. The top row of graphs shows that the majority of individual speakers in all three groups show a concave decreasing slope between the cues for voiced stops, that is, shorter negative VOTs are associated with lower onset *f*0. This is what would be expected if there was a trade-off between the two cues (§2.2). By

contrast, voiceless stops (bottom graphs) do not show any correlation between the two cues.

Quadratic regression analyses for vowel onset *f*0 and VOT covariation, performed on the group data, showed a significant negative correlation for on-target voiced stops in control speakers [β_2 = 1.60, $F_{(2,158)}$ = 5.86, p = .003] and CA speakers [β_2 = 2.01, $F_{(2,77)}$ = 20.91, p = .000], but not for AOS speakers (due primarily to one of the speakers). Voiceless stops did not show significant correlations between the two cues for any of the groups.

Figure 2. VOT and *f*0 correlation fitted with a regression function for each speaker for on-target voiced (top) and voiceless (bottom) stops by group.



Turning to the analysis of 'errors', that is, 'devoiced' stops (voiced stops identified as voiceless), only the AOS group produced these errors in sufficient numbers to merit analysis. Fig. 3 presents the *f*0-VOT covariance for this group. Regression lines are drawn for individual speakers. The figure shows a negative *f*0-VOT correlation for the majority of AOS speakers: for longer positive VOTs, onset *f*0 was lower, indicating compensatory behavior. Linear regression analyses revealed a significant negative covariation between the cues for this group, where for each millisecond of increase in VOT, *f*0 values decreased by 1.59 Hz (β_1 = -1.59, $F_{(1,25)}$ = 5.67, p = .025).

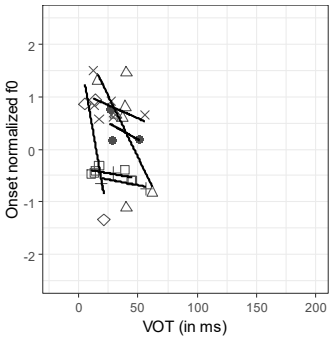


Figure 3. AOS group: VOT and *f*0 correlation fitted with a regression function for each speaker for 'devoiced' stops.

4. DISCUSSION

Our results indicate that speakers with apraxia and conduction aphasia (and also control speakers) exhibit covariation of acoustic cues in voiced stops. While the fine coordination of laryngeal and supralaryngeal gestures required for voiced stops is perturbed in the two clinical groups, most speakers — including healthy speakers— show a *negative* correlation between onset *f*0 and VOT, such that voiced stops with *short* prevoicing are produced with *lower* onset *f*0 (Fig. 2, top). Since shorter negative VOTs do not optimally cue voiced stops in Spanish, the lower *f*0 association in the three groups may be hypothesized to be the result of compensatory articulation. These results are in line with previously reported evidence for American English [12, 20]. They are not in accord, however, with those reported in [8], where no correlation was found between onset *f*0 and VOT within the voiced and voiceless categories in Spanish.

The predictions of the HSFC model receive only partial support in our data. Specifically, we expected to find compensatory articulation in voiced stops in apraxic speech. Examination of individual speakers (Fig. 2, top) shows that three out of six AOS speakers show steep negative correlations between the cues for voiced stops. Interestingly, AOS speakers also show significant negative slopes in devoiced stops (Fig. 3), that is, longer voicing lags are associated with lower onset *f*0 values. These results suggest that, aware that their motor plans and VOT values do not land within the intended phonological category (voiced), AOS speakers attempt to compensate by lowering *f*0. In sum, the AOS group tends to exhibit articulatory compensation in voiced stops and devoiced stops. On the other hand, against the predictions of the HSFC model, the CA group also exhibited compensatory articulation (i.e., negative slopes) in voiced stops. A word of caution is in order, however. Because the same negative VOT-*f*0 correlations are found in both the clinical and healthy groups, these patterns may be a property of cue covariation, and compensatory patterns may be speaker-dependent.

Finally, regarding the automatic vs. controlled (enhancing) nature of *f*0 variation associated with preceding consonant voicing, our results as well as our interpretation of compensatory articulation are compatible with a controlled account. Our findings are not in line with those reported in [13], where a positive VOT-*f*0 correlation is found for French and Italian, which is interpreted as supporting a biomechanical account (though the authors do not discard the possibility that speakers may enhance the effect).

5. REFERENCES

[1] Bang, H.-Y. 2017. The structure of multiple cues to stop categorization and its implications for sound change. PhD dissertation, McGill University, Canada.

[2] Blumstein, S. E., Cooper, W. E., Goodglass, H., Statlender, S., Gottlieb, J. 1980. Production deficits in aphasia: A voice-onset time analysis. *Brain and Language* 9(2), 153–170.

[3] Bohland, J.W., Bullock, D., Guenther, F.H. (2010). Neural Representations and Mechanisms for the Performance of Simple Speech Sequences. *Journal of Cognitive Neuroscience*, 22 (7), pp. 1504-1529.

[4] Bourguignon, N., Baum, S.R., Shiller, D.M. 2014. Lexical-perceptual integration influences sensorimotor adaptation in speech. *Frontiers in Human Neuroscience* 8(208).

[5] Buchwald, A., Miozzo, M. 2012. Phonological and motor errors in individuals with acquired impairment. *Journal of Speech, Language, and Hearing Research* 55(5), 1573–1586.

[6] Buckingham, H. W., Christman, S. 2008. Disorders of phonetics and phonology. In B. Stemmer, H. A. Whitaker (eds.), *Handbook of the Neuroscience of Language*. London: Academic Press, 127–136.

[7] Caudrelier, T., Schwartz, J.-L., Perrier, P., Gerber, S., Rochet-Capellan, A. (2018. Transfer of learning: What does it tell us about speech production units? *Journal of Speech and Hearing Research* 61(7), 1613–1625.

[8] Dmitrieva, O., Llanos, F., Shultz, A. A., Francis, A. L. 2015. Phonological status, not voice onset time, determines the acoustic realization of onset *f*0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics* 49, 77–95.

[9] Guenther, F.H. (2006) Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, 39, pp. 350-365.

[10] Hickok, G. 2012. Computational neuroanatomy of speech production. *Nature Reviews. Neuroscience* 13(2), 135–145.

[11] Katseff, S., Houde, J. F., Johnson, K. 2012. Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech* 55(2), 295–308.

[12] Khasanova, A., Cole, J., & Hasegawa-Johnson, M. 2014. Detecting articulatory compensation in acoustic data through linear regression modelling. In *Interspeech 2014* Singapore, 925–929.

[13] Kirby, J., Ladd, R. D. 2015. Stop voicing and f0 perturbations: Evidence from French and Italian. *Proceedings 18th ICPhS* Glasgow.

[14] Kirby, J., Ladd, R. D. 2016. Effects of obstruent voicing on vowel *F*0: Evidence from "true voicing" languages. *Journal of Acoustic Society of America* 140(1), 2400–2411.

[15] Labos, E., Zabala, K., Del Rio, M., Nespoulous, J.-L. 2005. *Protocolo Montréal – Toulouse - Buenos Aires de examen lingüístico de la Afasia MTBA05*. Buenos Aires: Ediciones Lenguaje y Cognición.

[16] Laganaro, M. 2015. Paraphasies phonémique et/ou phonétiques ? Des raisons et des difficultés de cette distinction. *Revue de Neuropsychologie* 7, 27–32.

[17] Niziolek, C. A., Guenther, F. H. 2013. Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience* 33(29), 12090 –12098.

[18] Perkell, J. S., Matthies, M. L., Svirsky, M. A., Jordan, M. I. 1995. Goal-based speech motor control: A theoretical framework and some preliminary data. *Journal of Phonetics* 23(1–2), 23–35.

[19] Perrier, P., Fuchs, S. 2015. Motor equivalence in speech production. In M. R. Redford (ed.), *The Handbook of Speech Production*. Oxford: Wiley-Blackwell, 225–247.

[20] Shultz, A., Francis, A. L., Llanos, F. 2012. Differential cue weighting in perception and production of consonant voicing. *Journal of Acoustic Society of America* 132(2). EL95-EL101.

[21] Vijayan, A., Gandour, J. 1995. On the notion of a "subtle phonetic deficit" in fluent/posterior aphasia. *Brain and Language* 48(2), 106–119.

[22] Ziegler, W. 2002. Psycholinguistic and motor theories of apraxia of speech. *Seminars in Speech and Language* 23, 231–243.

[1] This study was supported by grants FFI2017-84479-P from the Spanish Ministry of Science and Innovation and 2017SGR34 from the Generalitat de Catalunya and by grants by grants ANR-11-LABX-0036 (BLRI), ANR-11-IDEX-0001- (A*MIDEX).

[2] For three control speakers the beginning of the vowel after *voiceless* stops was devoiced and Praat did not find the *f*0. Thus Fig. 2 shows only data points and regression lines for three control speakers.