PERCEPTUAL ADAPTATION TO ALTERED CUE INFORMATIVENESS: DISTRIBUTIONAL, AUDITORY, AND LEXICAL FACTORS

Jiayin Gao[1,2,a)], James Kirby[2,b)]

[1]LPP (UMR7018, CNRS/Sorbonne Nouvelle),
[2]Institute for Phonetics and Speech Processing (IPS), LMU Munich
a) jiayin.gao@cnrs.fr, b) jkirby@phonetik.uni-muenchen.de

ABSTRACT

This study examines French-speaking listeners' perceptual adaptation to altered informativeness of acoustic cues signalling the voicing contrast. We attempt to condition a shift of perceptual attention from a primary cue to a secondary cue. Listeners were exposed to stimuli in which the distributional characteristics of either the primary or secondary cue were altered: either informativeness of F0 was increased, or that of VOT was decreased. When lexical feedback was given, increased informativeness of F0 led to listeners' increased use of high F0 in identifying voiceless plosives, whereas decreased informativeness of VOT led to increased use of low F0 in identifying voiced plosives. We suggest this asymmetry is due to general auditory mechanisms in the perception of F0 and VOT. Our findings highlight the importance of non-distributional characteristics such as auditory mechanisms and lexical knowledge in cue (re-)weighting, shedding light on how a secondary cue is phonologised in sound change.

Keywords: perceptual learning; perceptual adaptation; cue weighting; voicing; French

1. INTRODUCTION

Phonetic categories are signalled by multiple cues. For example, VOT (voice onset time) and F0 (fundamental frequency) (among other acoustic dimensions) signal the voicing contrast of plosives in many languages, with larger/positive VOTs associated with a higher F0 at the vowel onset (or onset F0) and smaller/negative VOTs with a lower onset F0 (e.g., [1]). But the relative weight of these cues differs. When the informativeness of one acoustic dimension is greater, listeners give a greater perceptual weight to this dimension [2, 3].

Following [4, 5, 6, 7], we define "informativeness" in distributional and probabilistic terms. When the distributions of two phonetic categories are more separated along one acoustic dimension, this dimension has greater informativeness than when their distributions are more overlapped (all else being equal). In languages like French or English, the VOT distributions of the voiced and voiceless categories are more separated than the F0 distributions. The VOT dimension is thus more informative, and listeners typically give a greater perceptual weight to VOT than onset F0 in categorising voicing [8, 9, 10].

Cue weighting is also dynamic. When the informativeness of a cue is increased, listeners rapidly shift their attention to this cue through exposure [4, 6, 11], a process known as "perceptual adaptation". But how is listeners' perception affected by a *decrease* in the informativeness of the primary cue? This question is important because in real speech, the primary cue cannot always be reliably recovered, which, in the long term, could lead to a cue re-weighting process. The current study tackles this question by comparing speech perception in two experimental conditions: (1) in which the informativeness of a secondary cue is increased, and (2) in which the informativeness of the primary cue is decreased. The rationale for the second condition is that although manipulation of the primary cue does not modify the informativeness of secondary cues *per se*, it impacts their *relative* informativeness. Our goal is to understand the conditions under which listeners may come to upweight a secondary cue.

We explore this issue in the context of the French voicing contrast. In Experiment 1, we examine the effect of increased informativeness of F0 (a secondary cue) on categorisation of voiced and voiceless plosives. In Experiment 2, we examine the effect of decreased informativeness of VOT (a primary cue) on the same categorisation task.

We also examine the influence of lexical feedback. Although some studies (e.g., [6, 12]) have shown that changes to distributional informativeness alone can lead to perceptual adaptation, others have found that perceptual learning is more robust when lexical information is provided [13, 14, 15, 16]. We therefore probe for a potential effect of lexical feedback by displaying or not the appropriate image.

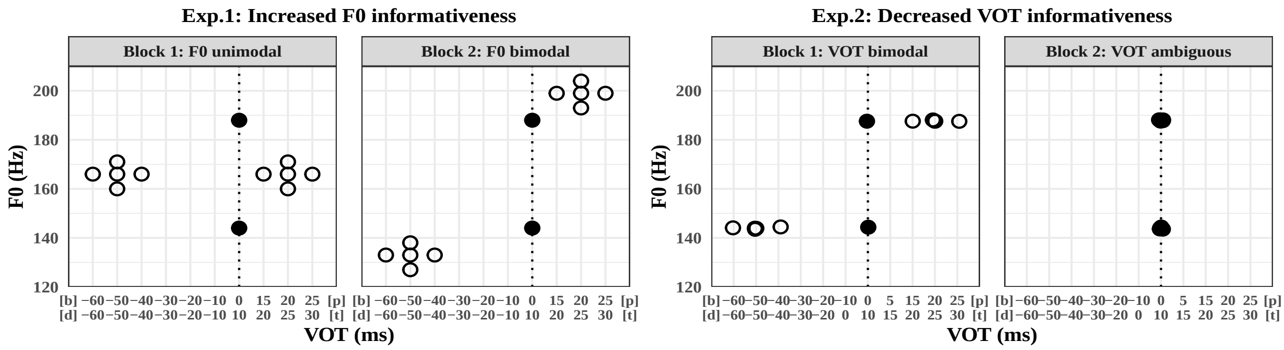

Figure 1: Stimuli of Experiments 1 (left) and 2 (right): each dot for a VOT-F0 association. Black dots are the test stimuli located at the ambiguous VOT step (dotted line). White dots are the exposure stimuli. For Experiment 2, jittered dots at -50, 20/25 ms (Block 1) and 0 ms (Block 2) represent higher numbers of stimulus trials.

2. EXPERIMENTS

2.1. Participants

120 native French-speaking listeners, recruited and paid through Prolific, participated remotely in two two-alternative-forced-choice (2AFC) experiments via PsyToolkit [17, 18]. 60 participated in Experiment 1 and the other 60 in Experiment 2. They were instructed to use earbuds or headphones and to perform the experiment in a quiet room. We discarded the data from one listener due to a high error rate (10 out of 30) in the training session.

2.2. Stimulus creation and protocol

A stimulus pool was created based on two minimal pairs, *pain* 'bread' vs. *bain* 'bath' and *thym* 'thyme' vs. *daim* 'deer', produced by a female native speaker of French. All four words contain the same rime [ɛ̃]. Continua along three parameters were created using re-synthesis. First, voice lead was manipulated from −60 to 0 ms with a step of 10 ms by reducing the original voice lead in *bain* or *daim* using cosine-function squared smoothing [19]. Second, voice lag was manipulated from 5 to 30 ms with a step of 5 ms by modifying the duration of the release within the original *pain* or *thym* using TD-PSOLA [20]. Third, onset F0 was manipulated from 127 to 204 Hz with a step of 5.5 Hz, again using TD-PSOLA.

For each experiment, a subset of stimuli were selected to form the distributional configurations of VOT (involving voice lead and voice lag) and onset F0 (Fig. 1), as detailed below. Participants performed the 2AFC task using their own personal computer. On each trial, the listener was presented with an auditory stimulus drawn either from the *pain–bain* pair or the *thym–daim* pair, and invited to choose the word they heard by clicking on one of the two pictures presented on the screen. As detailed below, the number of trials used to create an effect of exposure was decided following [6].

2.3. Experiment 1: Increased F0 informativeness

2.3.1. Procedure

Experiment 1 examined listeners' perceptual use of F0 when the informativeness of F0 was increased. In two blocks, listeners were exposed to a series of randomly ordered stimuli that fell in a given distributional configuration of VOT and F0, schematised in Fig. 1 (left), where each dot represents an association between VOT and onset F0 for a given stimulus. Block 1 was followed by Block 2.

In Block 1 (F0 unimodal), the distributions of voiced and voiceless categories were well separated along the VOT dimension but unimodal along the F0 dimension. That is, only VOT, but not F0, was informative. In Block 2 (F0 bimodal), the distributions of the two categories were equally well separated along the VOT dimension as in Block 1, but also separated along the F0 dimension. The F0 distance between the two categories was 65 Hz, above the post-obstruent F0 difference canonically observed in French. That is, the informativeness of VOT was unchanged, but that of F0 was increased.

The exposure stimuli, represented by the empty dots, fell within the distributional configurations described above. The two blocks also contained stimuli that were ambiguous for voicing identification, determined by the results of a baseline experiment conducted with a different listener group. They were the test stimuli, represented by the black dots. The task of the listener was the same for both exposure and test stimuli, but only their responses to the test stimuli were measured and compared across blocks.

In order to examine the effect of lexical feedback

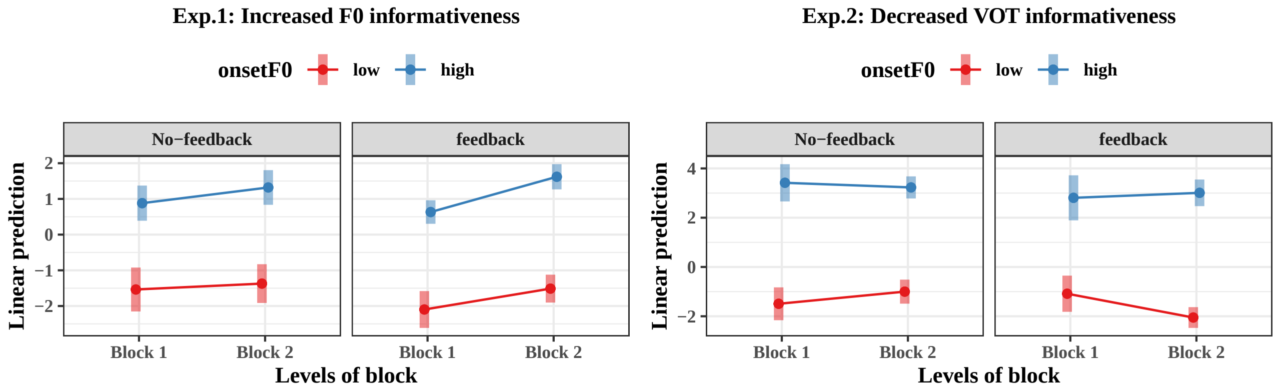

Figure 2: Interaction plots: voiceless response by F0, block, and feedback in the two experiments. Bars: 95% CIs.

associated with the F0 informativeness on listeners' performance, half of the listeners received feedback in Block 2 and the other half received no feedback. Feedback was provided by displaying the appropriate picture in the middle of the screen for 1.5 sec once the listener responded. Regardless of VOT, the pictures representing *bain* and *daim* were given as feedback to the auditory stimuli with a low onset F0 (from 127 to 144 Hz), and the pictures representing *pain* and *thym* were given as feedback to the auditory stimuli with a high onset F0 (from 188 to 204 Hz).

The no-feedback version contained 416 trials in total (12 exposure and test stimuli $\times$ 2 word pairs $\times$ 8 repetitions $\times$ 2 blocks + 32 attention checks). The feedback version contained 312 trials in total (12 exposure and test stimuli $\times$ 2 word pairs $\times$ 6 repetitions $\times$ 2 blocks + 24 attention checks).

2.3.2. Prediction

If an increase in the informativeness of onset F0 in distributional terms *alone* prompts listeners to upweight their use of onset F0 in voicing categorisation, we predict an increase in the use of F0 to the test stimuli in Block 2 compared to Block 1 (i.e., a greater use of high onset F0 in identifying voiceless plosives and/or a greater use of low onset F0 in identifying voiced plosives). If lexical feedback plays a role in addition to the distributional characteristics, we predict that such an increase will occur only or especially when lexical feedback is given.

2.3.3. Analysis and results

A generalised linear mixed model (GLMM), using the `lmerTest` package [21] in R [22], was fitted to listeners' response data to the test stimuli (i.e., the black dots in Fig. 1). The Helmert-coded predictors *Block* (1 vs. 2), *onsetF0* (high vs. low), *feedback* (no-feedback vs. feedback) and their three-way interaction were included. Random intercepts were included for participants, and by-participant random slopes for *Block*, *onsetF0*, and *feedback*. Interaction plots and pairwise comparisons were made using the `emmeans` package [23].

As shown in Fig. 2 (left), in the absence of lexical feedback, listeners' voiceless response rate increased in Block 2 as compared to Block 1 for high onset F0 only ($est. = 0.44, z = 2.45, p < .05$) but response rate did not differ between the two blocks for low onset F0 ($est. = 0.17, n.s.$). When lexical feedback was given, listeners' voiceless response rate increased in Block 2 compared to Block 1 for both F0 levels, but the increase was of a greater magnitude when onset F0 was high ($est. = 0.99, z = 4.87, p < .0001$) than when it was low ($est. = 0.59, z = 2.54, p < .05$). In sum, an increase in the informativeness of F0 led listeners to upweight the use of *high* onset F0 in identifying voiceless plosives. Lexical feedback strengthened this effect.

2.4. Experiment 2: Decreased VOT informativeness

2.4.1. Procedure

Experiment 2 examined listeners' perceptual use of F0 as the informativeness of VOT decreases, with two different distributional configurations, schematised in Fig. 1 (right). The test stimuli are the same as for Experiment 1. In Block 1 (VOT bimodal), the exposure stimuli are represented by the empty dots and the test stimuli by the black dots. Voiced and voiceless categories were well separated along both the VOT and F0 dimensions. Furthermore, VOT covaried with F0, with positive VOTs associated with a high onset F0 and negative VOTs with a low onset F0. That is, both VOT and F0 were informative. In Block 2 (VOT ambiguous), the test stimuli and the exposure stimuli (black dots) were the same. Stimuli were well separated along the F0 dimension but shared the same VOT step, which was ambiguous for voicing identification (i.e., VOT was distributionally uninformative).

As in Experiment 1, half the listeners received lexical feedback (display of the appropriate picture in Block 2), while the other half did not. Regardless of VOT, the pictures representing *bain* and *daim* were displayed as feedback accompanying auditory stimuli with a low onset F0 (144 Hz), and the pictures representing *pain* and *thym* were given as feedback to the auditory stimuli with a high onset F0 (188 Hz).

The no-feedback version contained 406 trials in total (12 exposure and test stimulus trials (Block 1) $\times$ 2 word pairs $\times$ 8 repetitions + 1 test stimulus (Block 2) $\times$ 2 word pairs $\times$ 90 repetitions + 34 attention checks). The feedback version contained 310 trials in total (12 exposure and test stimulus trials (Block 1) $\times$ 2 word pairs $\times$ 6 repetitions + 1 test stimulus (Block 2) $\times$ 2 word pairs $\times$ 70 repetitions + 26 attention checks).

2.4.2. Prediction

If the decrease in the informativeness of VOT in distributional terms *alone* upweights listeners' use of F0 in voicing categorisation, we predict an increase in the use of F0 to the test stimuli in Block 2 compared to Block 1. Again, if lexical feedback plays a role, we predict that such an increase will occur only or especially when lexical feedback is given.

2.4.3. Analysis and results

A GLMM with the same structure as for Experiment 1 was fitted to the response data to the test stimuli. As shown in Fig. 2 (right), when lexical feedback was *not* given, listeners' voiceless response rate was slightly increased in Block 2 as compared to Block 1 for low onset F0 ($est. = 0.49, z = 1.88$, $p = .061$), but did not differ between the two blocks for high onset F0 ($est. = -0.19, n.s.$). When lexical feedback was given, listeners' voiceless response rate was decreased in Block 2 as compared to Block 1 for low onset F0 ($est. = -0.97, z = -3.12$, $p < .005$), but did not differ for high onset F0 ($est. = 0.20, n.s.$).

In sum, while listeners did not respond to a decrease in the informativeness of VOT alone, the presence of lexical feedback led to an upweighted use of F0. Contrary to the increased use of high onset F0 in Experiment 1, here listeners increased their use of *low* onset F0 in identifying voiced plosives.

3. DISCUSSION

Our results first show that increased informativeness of F0 leads to a greater use of *high* onset F0 in identifying voiceless plosives, whereas a decreased informativeness of VOT leads to a greater use of *low* onset F0 in identifying voiced plosives, the latter being achieved only when lexical feedback is given. We hypothesise that general auditory plays a role in this asymmetry. In Experiment 1, while the raising and lowering of F0 is symmetrical in acoustic terms, listeners may be more sensitive to F0 raising at the auditory/attentional level [24]. In Experiment 2, the removal of prevoicing from voiced plosives may trigger a perceptual shift to low onset F0 as a replacement of the low-frequency energy at the vowel onset [25], giving rise to an increased perceptual weight of low onset F0. In contrast, the removal of voice lag from voiceless plosives does not affect the low-frequency energy at the vowel onset, hence F0 is not used as a rescue option. A similar asymmetry may then be expected in other languages where the voicing contrast is primarily signalled by the presence vs. absence of closure voicing such as Italian.

Our second finding, that listeners' increased use of high or low onset F0 only occurs or is strengthened when lexical feedback is given, is in line with other studies showing a crucial effect of lexical feedback in perceptual learning [13, 14, 15, 16]. Furthermore, it suggests that changes in distributional informativeness are processed by the listeners, but that for at least some kinds of contrasts or regions of phonetic space, distributional changes alone may not be sufficient to trigger a perceptual adaptation.

In sum, our study shows how a particular combination of distributional, auditory, and lexical factors contributes to the dynamic processing of VOT and F0 in a true voicing language. They suggest that F0 raising after voiceless plosives is perceptually more salient than F0 lowering after voiced plosives [26], mirroring the production pattern [27]. However, the exposure to devoiced tokens (when combined with listeners' implicit lexical knowledge) seems most favourable for increasing the perceptual weight of F0 lowering. Such laboratory findings may have implications for cue re-weighting in sound change (e.g., [28, 29, 30, 31]), pointing to perceptual mechanisms involved in the phonologisation of co-intrinsic F0 (aka tonogenesis) or a secondary cue in general.

4. ACKNOWLEDGEMENTS

This work was funded by ERC StG 758605 to J. Kirby. Thanks to M. Clayards, J. Schertz, J. Riverin-Coutlée and M. Gubian for helpful comments.

5. REFERENCES

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

[2] Holt, L. L., Lotto, A. J. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *JASA* 119(5), 3059–3071.

[3] Schertz, J., Clare, E. J. 2020. Phonetic cue weighting in perception and production. *WIREs Cognitive Science* 11(2), e1521.

[4] Clayards, M., Tanenhaus, M. K., Aslin, R. N., Jacobs, R. A. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108(3), 804–809.

[5] McMurray, B., Aslin, R. N., Toscano, J. C. 2009. Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science* 12(3), 369–378.

[6] Idemaru, K., Holt, L. L. 2011. Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance* 37(6), 1939–1956.

[7] Kirby, J. 2013. The role of probabilistic enhancement in phonologization. In: Yu, A. C. L. (ed), *Origins of Sound Change: Approaches to phonologization*. Oxford University Press, 228–246.

[8] Serniclaes, W. 1987. *Étude expérimentale de la perception du trait de voisement des occlusives du français*. PhD Thesis. Université Libre de Bruxelles.

[9] Abramson, A. S., Lisker, L. 1985. Relative power of cues: F0 shift versus voice timing. *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 25–33.

[10] Summerfield, Q., Haggard, M. 1974. Perceptual processing of multiple cues and contexts: Effects of following vowel upon stop consonant voicing. *Journal of Phonetics* 2(4), 279–295.

[11] Francis, A. L., Nusbaum, H. C. 2002. Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance* 28(2), 349–366.

[12] Kim, D., Clayards, M., Kong, E. J. 2020. Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics* 81, 100984.

[13] Norris, D., McQueen, J. M., Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology* 47(2), 204–238.

[14] Harmon, Z., Idemaru, K., Kapatsinski, V. 2019. Learning mechanisms in cue reweighting. *Cognition* 189, 76–88.

[15] Kraljic, T., Samuel, A. G. 2006. Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review* 13(2), 262–268.

[16] Goudbeek, M., Cutler, A., Smits, R. 2008. Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication* 50(2), 109–125.

[17] Stoet, G. 2010. PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods* 42(4), 1096–1104.

[18] Stoet, G. 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology* 44(1), 24–31.

[19] Mitterer, H. Generating fade-in and fade-outs (personal web page).

[20] Valbret, H., Moulines, E., Tubach, J.-P. 1992. Voice transformation using PSOLA technique. *Speech Communication* 11(2-3), 175–187.

[21] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13).

[22] R Core Team, 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

[23] Lenth, R. 2022. emmeans: Estimated marginal means, aka Least-squares means.

[24] Hsu, C. H., Evans, J. P., Lee, C. Y. 2015. Brain responses to spoken F0 changes: Is H special? *Journal of Phonetics* 51, 82–92.

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

[26] Gao, J., Yun, J., Arai, T. 2019. VOT-F0 coarticulation in Japanese: Production-biased or misparsing. *Procs. of the 19th ICPhS*, paper 619.1–5.

[27] Kirby, J., Ladd, D. R. 2016. Effects of obstruent voicing on vowel F0: evidence from "true voicing" languages. *JASA* 140(4), 2400–2411.

[28] Beddor, P. S. 2009. A coarticulatory path to sound change. *Language* 85(4), 785–821.

[29] Coetzee, A. W., Speeter, P., Shedden, K., Styler, W., Wissing, D. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66, 185–216.

[30] Kuang, J., Cui, A. 2018. Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics* 71, 194–214.

[31] Harrington, J., Kleber, F., Reubold, U., Schiel, F., Stevens, M. 2019. The phonetic basis of the origin and spread of sound change. In: Katz, W. F., Assmann, P. F. (eds), *The Routledge Handbook of Phonetics*. New York: Routledge.