Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking

ONLINE SUPPLEMENTAL MATERIALS

Efthymia C. Kapnoula
Dept. of Psychological and Brain Sciences
DeLTA Center
University of Iowa
Basque Center on Cognition, Brain and Language

and

Bob McMurray
Dept. of Psychological and Brain Sciences
Dept. of Communication Sciences and Disorders
Dept. of Linguistics
DeLTA Center
University of Iowa

S1. VAS task results: Hierarchical regression results

Table S1.1 Results of hierarchical regression analyses predicting *raw (left side) and residualized (right side) VAS slopes as a function of place of articulation (PoA) and secondary cue use (*θ *angle).*

		DV: VAS slope			DV: Residualized VAS slope		
Step	Variable	*β*	ΔR^2	*p*	*β*	ΔR^2	*p*
1	PoA	-0.328	0.107	< .001	-0.023	<0.001	.792
2	PoA	-0.150		.156	0.181		
2	θ angle	0.277		.010	0.318		
			0.045	< .001		0.060	.018

S2. Identification (mouse-clicking) results in the lexical gradiency VWP task

While the primary measure in the VWP Gradiency task was based on the fixations, we also examined the mouse-clicking responses to establish the expected patterns of effects. This analysis examined the likelihood of participants clicking on the picture of the unvoiced word in a pair (*likelihood of unvoiced response* or *LUR*). We fitted a logistic mixed effects model using the *glmer* function in R, with VOT, F_0, and place of articulation (PoA), and their interactions as fixed effects. VOT and F_0 were centered and PoA was effect-coded (alveolar = 1; labial = -1). The random effects included a random slope of VOT, PoA, and their interaction for subjects (see Equation 1, Table S2.1 for results).

LUR ~ VOT*F0*PoA + (VOT*PoA |subject), family = binomial (1)

There was a significant main effect of VOT, $p < .001$, with greater VOT predicting higher LUR, as expected. There was also a main effect of F_0, $p < .001$, in the expected direction; higher F_0 predicted higher likelihood of unvoiced classification. PoA was also significant, $p < .001$, with participants being more likely to give an unvoiced response for labial stimuli. The VOT × F_0 interaction was significant, $p < .001$, as was the three-way interaction, $p < .001$, showing a stronger F_0 effect for labial stimuli with low VOTs (see Figure S2.1). Because of the significant effects of secondary variables (PoA and F_0), we did not collapse across these in subsequent analyses of the fixations.

Table S2.1 *Likelihood of unvoiced response as predicted by VOT, F_0, and place of articulation*

	Estimate	*S.E.*	*z*	*p*
(Intercept)	1.252	.092	13.623	< .001
VOT	1.927	.055	35.156	< .001
F_0	.243	.056	4.321	< .001
PoA	-1.029	.065	-15.930	< .001
VOT*F_0	-198	.042	-4.741	< .001
VOT*PoA	.023	.036	.651	.515
F_0*PoA	.002	.056	.028	.978
VOT*F_0*PoA	.154	.042	3.688	< .001

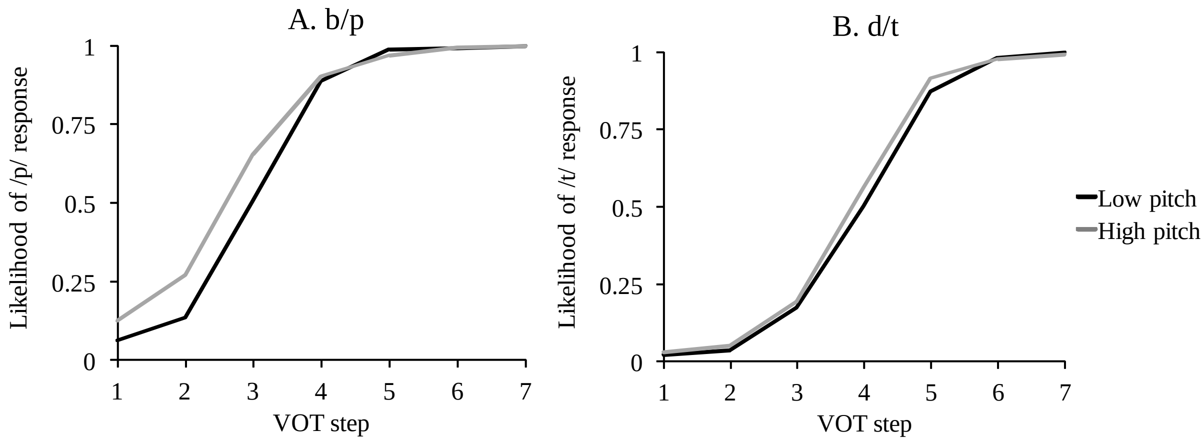

Figure S2.1. Likelihood of "unvoiced" response as a function of VOT and F_0 (pitch) for the A) b/p continuum and B) for the d/t continuum.

S3. Computation of rVOT

Analysis of the VWP and P3 data needed to account for variation in listeners' own category boundaries to ensure that the gradient effects were truly within-category. For example, a VOT of 20 ms might be a /b/ for some subjects, but a /p/ for others. Consequently, even if there was a perfectly discrete boundary (that varied between subjects), one could see enhanced competitor fixations at 20 ms, because some subjects perceived this as a /p/, while others heard it as a /b/. The same averaging logic also applies to items; for example, *beach/peach* could have a different boundary than *bowl/pole* due to coarticulation. To account for this, we used two techniques based on McMurray et al. (2016).

First, we computed a measure of *relative VOT* (rVOT) that reflected the category boundary for that subject and continuum. Here, -1 was one step to the left of the boundary – regardless of where that subject's boundary is located, -2 was two steps (etc). To accomplish this, we fit a four-parameter logistic to each participant's identification curve (based on the mouse click response). We then used the crossover parameter as an estimate of category boundary. We did this for each place of articulation and F_0 value separately, yielding four crossovers per participant. Fits were performed using a constrained gradient descent method implemented in Matlab (McMurray, 2017). We then did this for each continuum (collapsed over subject[1]). Next, we computed the estimated crossover on each trial by first computing the deviation of each item-specific crossover from the average crossover; next we added that deviation to the subject-specific crossover. As a result for each participant, we had an estimated category boundary that accounted for the effects of place of articulation, F_0, and item (see McMurray et al., 2016, for a similar procedure). We used this boundary estimate to calculate the distance between the participant's category boundary and the actual VOT step on each trial, (henceforth, *relative VOT* or *rVOT*; see also Kapnoula et al., 2021; McMurray et al., 2009). For example, given a crossover of 4.3, VOT step 6 would be "unvoiced" with an rVOT of 1.7, whereas VOT step 2 would be "voiced" with an rVOT of -2.3.

Second in addition to this boundary adjustment, our main analysis only included trials in which the participant's response matched their predicted response based on their estimated boundary (e.g., for an rVOT of -1, they should have selected the voiced response; for an rVOT of

[1] There were insufficient data in this shortened version of the paradigm to do this within each continuum within each subject (e.g., per condition / participant).

+2, the unvoiced). Thus, any differences in eye-movements would reflect within-category sensitivity.

S4. Detailed results for lexical gradiency VWP task

The analysis of the lexical gradiency VWP used two models. The first model is reported in Equation (2) below and statistical results are in Table S4.1.

Comp-Filler ~ rVOT*VASslope + PoA*Voicing + (1|subject) + (rVOT|item) (2)

Table S4.1. *Difference between competitor and filler looks (Comp-Filler) as predicted by rVOT, raw VAS slope, place of articulation (PoA), and voicing of target*

	Estimate	*S.E.*	*d.f.*	*t*	*p*
(Intercept)	0.019	0.006	8	3.480	.008
rVOT	0.008	0.001	9	6.621	<.001
VASslope	-0.005	0.003	251	-1.550	.122
PoA	-0.002	0.005	9	-0.352	.734
Voicing	-0.004	0.001	3233	-4.852	<.001
rVOT*VASslope	-0.004	0.002	5871	-2.249	.025
PoA*Voicing	-0.006	0.001	18550	-7.372	<.001

Note: Results mentioned in the MS are highlighted

We next replicated this analysis using VAS slope after it had been residualized for the performance in the visual VAS task (Equation 3, Table S4.2).

Comp-Filler ~ rVOT*ResVASslope + PoA*Voicing + (1|subject) + (rVOT|item) (3)

Table S4.2 *Difference between competitor and filler looks (Comp-Filler) as predicted by rVOT, residualized VAS slope, place of articulation (PoA), and voicing of target*

	Estimate	*S.E.*	*df*	*t*	*p*
(Intercept)	0.019	0.006	8.073	3.460	.008
rVOT	0.008	0.001	9.566	6.297	<.001
res.VASslope	-0.002	0.001	254.8	-1.582	.115
PoA	-0.002	0.005	8.429	-0.374	.717
Voicing	-0.004	0.001	3736	-4.808	<.001
rVOT*res.VASslope	-0.001	0.001	21420	-1.806	.071
PoA*Voicing	-0.006	0.001	19060	-7.377	<.001

S5. Phoneme identification results in the EEG/ERP task

The first goal of the accuracy analysis was to validate the task (e.g., to ensure that listeners were complying with the task, the stimuli were intelligible). Given that participants vary in VOT boundary, we used place of articulation as our criterion for accuracy; if they were monitoring for a /b/ and responded “target” for anything in the d/t continuum, that response was considered incorrect. By this measure, average accuracy was high (*M* = 98.6%, *SD* = .59%) and participants responded promptly (*M* = 387[2] ms, *SD* = 346 ms).

Next, we examined the proportion of target responses as a function of VOT and F_0. As expected, participants were affected by both VOT and F_0; stimuli with lower VOT/F_0 values were more likely to be categorized as voiced (and vice versa; see Figure S5.1).

To test this statistically, we re-coded VOT and F_0 in terms of their distance from the target (e.g., VOT step 7 was recoded as 6, when the target was voiced, and as 2, when the target

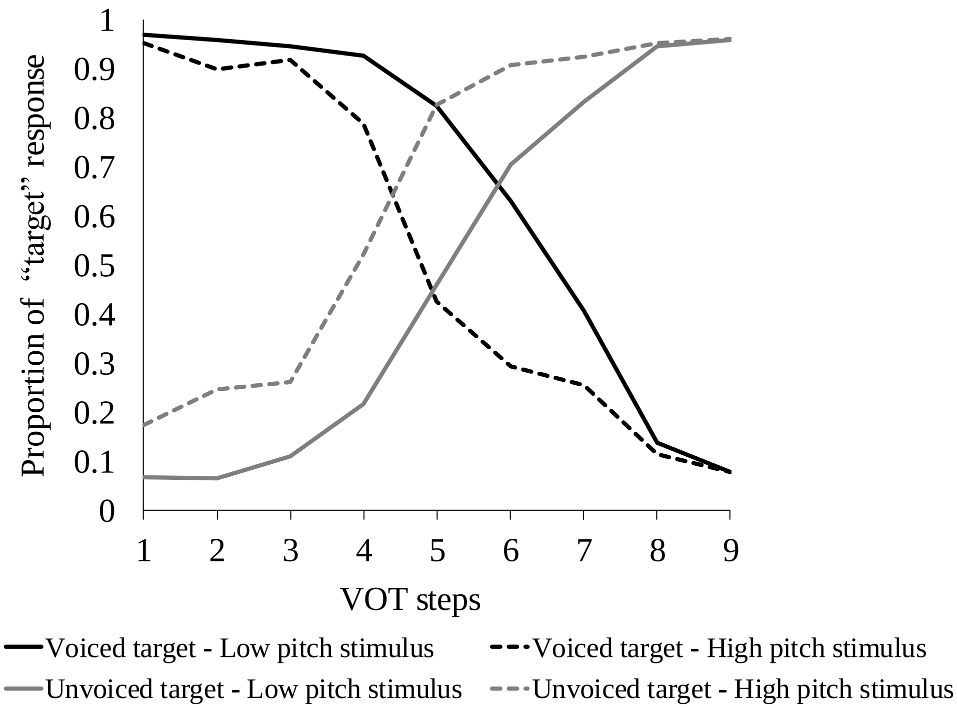

Figure S5.1. Proportion of “target” response as a function of voicing of target and VOT and F_0 of the stimulus.

---

[2] As mentioned in Methods, RT was calculated starting at the onset of the prompt (i.e., they were not allowed to respond before). The prompt appeared ~200 ms after the end of the word.

was unvoiced, whereas F_0 was recoded as matching/mismatching with the target). We entered these recoded VOT and F_0 variables (centered) and their interaction as fixed effects in a logistic mixed effects model. Place of articulation (PoA; effect-coded: alveolar = 1; labial = -1), target voicing (effect-coded: "voiced" = -1; "unvoiced" = 1), and their interaction were entered as covariates. The dependent variable was whether the participant made a "target" (coded as 1) or "other" response (coded as 0). The random effects structure included random intercepts and random VOT and F_0 slopes (and their interaction) for subjects (Equation 6, results in Table S5.1).

P(Target) ~ VOTdist * F_0dist + PoA*Voicing PoA + (VOTdist*F_0dist|subject), family = binomial (6)

Table S5.1 *Probability of target response as predicted by distance from target, in VOT and F_0, place of articulation, and voicing of target*

	Estimate	*S.E.*	*z*(67259)	*p*
(Intercept)	0.449	0.068	6.557	.000
VOTdist	-1.029	0.046	-22.255	< .001
F_0dist	-0.334	0.021	-16.086	< .001
PoA	0.209	0.012	17.667	< .001
Voicing	0.446	0.012	36.968	< .001
VOTdist*F_0dist	-0.001	0.005	-.174	.862
PoA*Voicing	-0.570	0.012	-46.483	< .001

As shown in Table S5.1, VOT distance was significant, $p < .001$, as was F_0 compatibility, $p < .001$. That is, greater VOT/F_0 distance from the target predicted lower probability of a "target" response, as expected. The VOT × F_0 interaction was not significant. Thus, overall, participants performed the ERP task as expected.

S6. Detailed P3 results

This analysis examined the P3 (average voltage within a time-window 400 to 730 ms post stimulus onset) as a function of stimulus prototypically (i.e., how target-like it was). Fixed effects included distance from the target (in VOT and F_0), a factor reflecting whether the stimulus and the target matched in place of articulation (PoA; coded as 1 for Match and -1 for Mismatch), and all interactions. Voicing of the target was added as a covariate. Data from 5 channels were included (Pz, P7, P3, CP2, and CP1; see main text Figure 4A) and channel was treated as a random effect (see Equation 7, Table S6.1 for results).

$$P3_{amp} \sim VOTdist * F_0dist * Match + TargetVoice + (VOTdist*F_0dist|subject) + (1|channel) \quad (7)$$

Table S6.1 *P3 amplitude as predicted by distance from the target in VOT steps (VOTdist) and F_0 steps (F_0dist), relevancy of the stimulus for the task (Match), and target voicing*

	Estimate	*S.E.*	*d.f.*	*t*	*p*
(Intercept)	-0.240	0.142	62	-1.697	.095
VOTdist	-0.025	0.006	66	-4.062	<.001
F_0dist	-0.083	0.026	66	-3.169	.002
Match	0.185	0.007	47350	26.390	<.001
TargetVoice	-0.005	0.007	47350	-0.739	.460
VOTdist×F_0dist	0.005	0.012	67	0.456	.650
VOTdist×Match	-0.024	0.003	47350	-8.703	<.001
F_0dist×Match	-0.012	0.014	47350	-0.859	.390
VOTdist×Match	0.007	0.005	47350	1.302	.193

Note: Results mentioned in the MS are highlighted

In the next analysis VAS slope was added to the fixed effects to ask whether individuals with different VAS categorization patterns showed differences in the link between VOT and N1 (see Equation 8, Table S6.2 for results).

$$P3_{amp} \sim VOTdist * F_0dist * VASslope*Match + TargetVoice + (VOTdist*F_0dist|subject) + (1|channel) \quad (8)$$

Table S6.2 *P3 amplitude as predicted by distance from the target in VOT steps (VOTdist) and F_0 steps (F_0dist), relevancy of the stimulus for the task (Match), and target voicing*

	Estimate	*S.E.*	*d.f.*	*t*	*p*
(Intercept)	-0.240	0.141	61	-1.706	.093
VOTdist	-0.026	0.006	65	-4.139	<.001
F_0dist	-0.086	0.026	64	-3.273	.002
VASslope	0.184	0.038	45040	4.824	<.001
Match	0.183	0.007	45560	25.727	<.001
TargetVoice	-0.008	0.007	45570	-1.088	.277
VOTdist×F_0dist	0.004	0.012	67	0.374	.710
VOTdist×VASslope	0.047	0.012	772	3.815	<.001
F_0dist×VASslope	0.037	0.060	464	0.612	.541
VOTdist×Match	-0.023	0.003	45570	-8.333	<.001
F_0dist×Match	-0.010	0.014	45570	-0.722	.470
VASslope×Match	-0.058	0.023	45560	-2.591	.010
VOTdist×F_0dist×VASslope	-0.024	0.024	684	-0.999	.318
VOTdist×F_0dist×Match	0.009	0.006	45570	1.574	.115
VOTdist×VASslope×Match	0.028	0.009	45560	3.249	.001
F_0dist×VASslope×Match	0.056	0.045	45560	1.244	.213
VOTdist×F_0dist×VASslope×Match	0.044	0.017	45560	2.499	.012

Note: Results mentioned in the MS are highlighted

Next, this analysis was replicated using residualized VAS slope (see Equation 9, Table S6.3 for results).

$$P3_{amp} \sim VOTdist * F_0dist * res.VASslope * Match + TargetVoice + (VOTdist * F_0dist|subject) + (1|channel) \quad (9)$$

Table S6.3 *P3 amplitude as predicted by distance from the target in VOT steps (VOTdist) and F_0 steps (F_0dist), relevancy of the stimulus for the task (Match), and target voicing*

	Estimate	*S.E.*	*d.f.*	*t*	*p*
(Intercept)	-0.238	0.140	61	-1.702	.094
VOTdist	-0.025	0.006	65	-4.118	<.001
F_0dist	-0.085	0.026	64	-3.245	.002
res.VASslope	-0.016	0.014	44420	-1.143	.253
Match	0.183	0.007	45560	25.714	<.001
TargetVoice	-0.008	0.007	45570	-1.091	.275
VOTdist×F_0dist	0.004	0.012	67	0.355	.724
VOTdist×res.VASslope	0.008	0.004	535	1.925	.055
F_0dist×res.VASslope	0.001	0.021	342	0.032	.975
VOTdist×Match	-0.023	0.003	45570	-8.331	<.001
F_0dist×Match	-0.010	0.014	45570	-0.729	.466
VASslope×Match	-0.018	0.007	45560	-2.351	.019
VOTdist×F_0dist×res.VASslope	-0.001	0.008	490	-0.080	.937
VOTdist×F_0dist×Match	0.009	0.006	45570	1.586	.113
VOTdist×res.VASslope ×Match	0.008	0.003	45570	2.826	.005
F_0dist×res.VASslope ×Match	0.026	0.015	45570	1.773	.076
VOTdist×F_0dist×res.VASslope ×Match	0.015	0.006	45570	2.621	.009

Note: Results mentioned in the MS are highlighted

S7. Detailed results for spatial Stroop task

Table S7.1 shows results of a hierarchical regression examining the degree to which spatial Stroop performance predicts gradiency in the VAS task. Each model was done twice with raw VAS slope (left columns) and VAS slope after revisualization for the visual VAS measure (right columns).

Table S7.1. *Raw and residualized VAS slopes predicted by spatial Stroop measures*

		DV: VAS slope			DV: Residualized VAS slope		
Step	Variable	*β*	*ΔR*^{2}	*p*	*β*	*ΔR*^{2}	*p*
1	Average accuracy	-0.076		.581	-0.026		.738
1	Average RT	-0.040		.773	0.001		.594
			0.005	.857		0.004	.863
2	Average accuracy	-0.028		.830	0.005		
2	Average RT	-0.120		.397	-0.098		
2	Congruency effect	0.367		.004	0.395		
			0.114	.004		0.132	.002

S8. Complete results in the lexical inhibition VWP task

Mouse-Click Accuracy and RT

Prior to the analyses, accuracy percentages were logit-transformed and RTs were log-transformed to normalize the positive-skewed distribution of the data. A one-way repeated measures ANOVA showed no effect of splicing on accuracy $F < 1$, which was not surprising given that responses were at ceiling (matching-splice: *M* = 99.6%, *SD* = 1.4%; nonword-splice: *M* = 99.6%, *SD* = 1.3%; word-splice: *M* = 99.6%, *SD* = 1.3%). A similar analysis of variance on RT revealed a significant splicing effect, $F(2,140) = 63.667$, $\eta^2 = .476$, $p<.001$.

Post-hoc comparisons (Bonferroni-corrected) showed that participants were slower in the competitor word-splice (*M* = 1272 ms, *SD* = 128 ms) compared to both the matching-splice (*M* = 1145 ms, *SD* = 115 ms), $F(1,70) = 123.55$, $\eta^2 = .638$, $p < .001$, and the control/nonword-splice (*M* = 1231 ms, *SD* = 125) condition, $F(1,70) = 11.12$, $\eta^2 = .137$, $p < .001$. The splicing effect on RT offers preliminary evidence that competitor activation in the word-splice condition slowed down target activation.

Table S8.1. *Subphonemic splicing effect on accuracy and reaction times*

Predictor	Sum of Squares	*d.f.*	Mean Square	*F*	*p*	partial η
	Dependent variable: Accuracy					
Splicing	0.056	2	.028	.155	.856	.00
Error	25.334	140	.181			
	Dependent variable: Reaction Time					
Splicing	0.078	2	.039	63.667	<.001	.47
Error	0.085	140	.001			

Table S8.2. *Splice effect on RT: Post-hoc comparisons (Bonferroni-corrected)*

Comparison						
Condition 1	*Condition 2*	*Mean Difference*	*S.E.*	*F*	*p*	partial η
Matching-splice	- Word-splice	-.046	.004	123.549	<.001	.63
Nonword-splice	- Word-splice	-.014	.004	11.120	.001	.13

Fixations

Fixations to the target (area under the curve 600 to 1600) were analyzed with a linear mixed effects model with Splice condition and VAS slope as fixed effects. Splice condition was coded as two contrasts, one contrasting the matching-splice and nonword-splice conditions, and one capturing the nonword-splice vs. word-splice condition (the effect of lexical inhibition). See Equation 10, and Table S8.3 for complete results).

$$\text{Target Looks} \sim ([ne_t t \text{ vs. } ne_p t] + [ne_p t \text{ vs. } ne_{ck} t]) * \text{res.VAS slope} + (1|\text{subject}) + (1|\text{item}) \quad (10)$$

This analysis was replicated using residualized VAS slope (Equation 11, Table S8.4)

$$\text{Target Looks} \sim ([ne_t t \text{ vs. } ne_p t] + [ne_p t \text{ vs. } ne_{ck} t]) * \text{res.VAS slope} + (1|\text{subject}) + (1|\text{item}) \quad (11)$$

Table S8.3. *Proportion of looks to the target as predicted by splice condition and VAS slope*

	Estimate	*S.E.*	*df*	*t*	*p*
(Intercept)	2.512	0.193	58.947	13.034	< .001
$ne_t t$ vs $ne_p t$	-1.055	0.136	4685.085	-7.779	<.001
$ne_p t$ vs $ne_{ck} t$	-0.678	0.136	4685.112	-5.002	<.001
VASslope	0.857	0.555	54.985	1.543	.128
[$ne_t t$ vs $ne_p t$] × VASslope	-0.470	0.569	4685.066	-0.827	.408
[$ne_p t$ vs $ne_{ck} t$] × VASslope	0.085	0.568	4685.064	0.150	.881

Note: Results mentioned in the MS are highlighted

Table S8.4. *Proportion of looks to the target as predicted by splice condition and residualized VAS slope*

	Estimate	*S.E.*	*df*	*t*	*p*
(Intercept)	2.512	0.193	59.238	13.001	< .001
$ne_t t$ vs $ne_p t$	-1.055	0.136	4685.083	-7.779	<.001
$ne_p t$ vs $ne_{ck} t$	-0.678	0.136	4685.110	-5.001	<.001
res.VASslope	0.182	0.137	54.988	1.331	.189
[$ne_t t$ vs $ne_p t$] × VASslope	-0.103	0.139	4685.076	-0.740	.459
[$ne_p t$ vs $ne_{ck} t$] × VASslope	-0.037	0.139	4685.062	-0.269	.788

S9. Detailed N1 results

S9.1 Modal effects

This analysis examined the N1 (average voltage within a time-window 115 to 170 ms post stimulus onset) as a function of VOT. Our goal was to replicate the analysis of Toscano et al. (2010) and extend it by adding F_0 which was not manipulated in that study. Fixed effects included VOT, F_0, and their interaction as fixed effects. Place of articulation of the stimulus and the target (PoA; alveolar: 1, labial: -1) and voicing of target (unvoiced: 1, voiced: -1), and their interactions were added as covariates. Data from 12 channels were included (Cz, CP1, CP2, C3, C4, FC2, FC1, CP5, CP6, Fz, FC5, and FC6; see main text Figure 7B,C) and channel was treated as a random effect (see Equation 12, Table S9.1 for results).

$$N1_{amp} \sim VOT*F_0 + PoA*TargetVoice*TargetPoA + (VOT*F_0|subject) + (VOT|channel) \quad (12)$$

Table S9.1. *N1 amplitude as predicted by VOT, F_0, place of articulation (PoA), target voicing, and target PoA*

	Estimate	*S.E.*	*d.f.*	*t*	*p*
(Intercept)	-1.206	0.269	37	-4.476	<.001
VOT	0.114	0.014	43	8.168	<.001
F_0	0.310	0.050	67	6.241	<.001
PoA	-0.089	0.007	114500	-12.029	<.001
TargetVoice	0.034	0.007	114500	4.606	<.001
TargetPoA	0.156	0.007	114500	21.143	<.001
VOT×F_0	-0.040	0.017	67	-2.314	.024
PoA×TargetVoice	0.013	0.007	114500	1.737	.082
PoA×TargetPoA	-0.139	0.007	114500	-18.787	<.001
TargetVoice×TargetPoA	-0.227	0.007	114500	-30.742	<.001
PoA×TargetVoice×TargetPoA	0.005	0.007	114500	0.622	.534

Note: Results mentioned in the MS are highlighted

S9.2 Individual differences

In the subsequent model, VAS slope (centered) was added to the fixed effects to test whether individuals with different VAS categorization patterns showed differences in the link between VOT and N1 (see Equation 13, Table S9.2 for results).

$$N1_{amp} \sim VOT*F_0*VASslope + PoA*TargetVoice*TargetPoA + (VOT*F0|subject) + (VOT|channel) \quad (13)$$

Table S9.2. *N1 amplitude as predicted by VOT, F_0, VAS slope, place of articulation (PoA), target voicing, and target PoA*

	Estimate	*S.E.*	*d.f.*	*t*	*p*
(Intercept)	-1.204	0.270	36	-4.453	<.001
VOT	0.111	0.014	41	8.062	<.001
F_0	0.299	0.051	66	5.839	<.001
VASslope	-0.159	0.049	107200	-3.282	.001
PoA	-0.113	0.009	110100	-12.248	<.001
TargetVoice	0.033	0.008	110200	4.355	<.001
TargetPoA	0.155	0.008	110200	20.474	<.001
VOT×F_0	-0.044	0.017	65.740	-2.528	.014
VOT×VASslope	-0.066	0.015	3456	-4.494	<.001
F_0×VASslope	-0.125	0.075	2978	-1.666	.096
PoA×TargetVoice	0.007	0.008	110200	0.948	.343
PoA×TargetPoA	-0.142	0.008	110200	-18.762	<.001
TargetVoice×TargetPoA	-0.231	0.008	110200	-30.517	<.001
VOT×F_0×VASslope	0.081	0.028	1892	2.833	<.001
PoA×TargetVoice×TargetPoA	0.001	0.008	110200	0.111	0.912

Note: Results mentioned in the MS are highlighted

This analysis was replicated using residualized VAS slope (see Equation 14, Table S9.3 for results).

$$N1_{amp} \sim VOT*F_0*res.VASslope + PoA*TargetVoice*TargetPoA + (VOT*F_0|subject) + (VOT|channel) \quad (14)$$

Table S9.3. *N1 amplitude as predicted by VOT, F_0, res.VAS slope, place of articulation (PoA), target voicing, and target PoA*

	Estimate	*S.E.*	*d.f.*	*t*	*p*
(Intercept)	-1.205	0.270	36	-4.457	<.001
VOT	0.110	0.014	41	8.024	<.001
F_0	0.298	0.051	66	5.901	<.001
res.VASslope	-0.059	0.015	107700	-4.023	<.001
PoA	-0.098	0.008	110200	-12.676	<.001
TargetVoice	0.033	0.008	110200	4.362	<.001
TargetPoA	0.155	0.008	110200	20.481	<.001
VOT×F_0	-0.043	0.018	65	-2.454	.017
VOT×VASslope	-0.010	0.005	2204	-1.884	.060
F_0×VASslope	-0.009	0.027	1847	-0.328	.743
PoA×TargetVoice	0.007	0.008	110200	0.940	.347
PoA×TargetPoA	-0.142	0.008	110200	-18.771	<.001
TargetVoice×TargetPoA	-0.231	0.008	110200	-30.508	<.001
VOT×F_0×res.VASslope	0.046	0.010	1281	4.566	.000
PoA×TargetVoice×TargetPoA	0.001	0.008	110200	0.102	.919

S9.3 Testing for a categorical component in some listeners

After establishing a basic moderation of the VOT effect by VAS gradiency we next used a series of models to ask whether less gradient listeners had a categorical component to their N1s. To do this, A *binary* variable reflecting stimulus identity (voiced/unvoiced) was entered in the models used above to ask whether a step-like function was simultaneously present along with the linear effect of VOT for some listeners. This binary variable (stepVOT) was set to -1 for VOTs below each subject's individual boundary (computed from the response functions) and +1 for VOTs above it. The model is given in Equation 15 (see Table 9.4 for results).

$$N1_{amp} \sim VOT*F_0*VASslope + stepVOT*VASslope + PoA*Voicing + (VOT*F_0|subject) + (VOT|channel) \quad (15)$$

Table S9.4. *N1 amplitude as a function of VOT, F_0, VAS slope, stepVOT, place of articulation (PoA), target voicing, and target PoA*

	Estimate	*S.E.*	*d.f.*	*t*	*p*
(Intercept)	-1.213	0.270	36	-4.492	<.001
VOT	0.094	0.014	49	6.546	<.001
F_0	0.299	0.051	66	5.838	<.001
VASslope	-0.247	0.050	107300	-4.979	<.001
stepVOT	0.053	0.014	109600	3.772	<.001
PoA	-0.117	0.009	110100	-12.414	<.001
TargetVoice	0.033	0.008	110200	4.350	<.001
TargetPoA	0.155	0.008	110200	20.477	<.001
VOT×F_0	-0.044	0.017	66	-2.530	.014
VOT×VASslope	-0.179	0.020	10830	-9.167	<.001
F_0×VASslope	-0.125	0.075	2979	-1.669	.095
VASslope×stepVOT	0.366	0.043	110200	8.534	<.001
PoA×TargetVoice	0.007	0.008	110200	0.955	.340
PoA×TargetPoA	-0.142	0.008	110200	-18.766	<.001
TargetVoice×TargetPoA	-0.231	0.008	110200	-30.531	<.001
VOT×F_0×VASslope	0.080	0.028	1895	2.825	.005
PoA×TargetVoice×TargetPoA	0.001	0.008	110200	0.122	.903

Note: Results mentioned in the MS are highlighted

Again, this analysis was replicated using residualized VAS slope (see Equation 16, Table S9.6 for results).

$$N1_{amp} \sim VOT*F_0*res.VASslope + stepVOT*res.VASslope + PoA*TargetVoice*TargetPoA + (VOT*F_0|subject) + (VOT|channel) \quad (16)$$

Table S9.5. *N1 amplitude as predicted by VOT, F_0, res.VAS slope, stepVOT, place of articulation (PoA), target voicing, and target PoA*

	Estimate	*S.E.*	*d.f.*	*t*	*p*
(Intercept)	-1.218	0.270	36	-4.509	<.001
VOT	0.095	0.014	49	6.632	<.001
F_0	0.298	0.051	66	5.900	<.001
res.VASslope	-0.087	0.015	107800	-5.708	<.001
stepVOT	0.049	0.014	109800	3.523	<.001
PoA	-0.102	0.008	110200	-12.835	<.001
TargetVoice	0.033	0.008	110200	4.363	<.001
TargetPoA	0.155	0.008	110200	20.484	<.001
VOT×F_0	-0.043	0.018	65	-2.455	.017
VOT×res.VASslope	-0.042	0.007	6244	-6.250	<.001
F_0× res.VASslope	-0.009	0.027	1848	-0.331	.741
res.VASslope×stepVOT	0.106	0.014	110200	7.495	<.001
PoA×TargetVoice	0.007	0.008	110200	0.946	.344
PoA×TargetPoA	-0.142	0.008	110200	-18.776	<.001
TargetVoice×TargetPoA	-0.231	0.008	110200	-30.516	<.001
VOT×F_0× res.VASslope	0.046	0.010	1283	4.561	<.001
PoA×TargetVoice×TargetPoA	0.001	0.008	110200	0.108	.914

In order to conduct post-hoc analyses, we split the data by participants' VAS slope (average across places of articulation) and repeated the analyses without VAS slope in the fixed effects (see Equation 17). This was done for both raw VAS slope (Table S9.6) and residualized (Table S9.7).

$$N1_{amp} \sim VOT*F_0 + stepVOT + PoA*TargetVoice*TargetPoA + (VOT*F_0|subject) + (VOT|channel) \quad (17)$$

Table S9.6. *N1 amplitude as predicted by VOT, F_0, VAS slope, stepVOT, place of articulation (PoA), target voicing, and target PoA for gradient participants only*

	Estimate	*S.E.*	*d.f.*	*t*	*p*
(Intercept)	-1.038	0.343	43	-3.031	.004
VOT	0.126	0.020	48	6.333	<.001
F_0	0.274	0.071	33	3.842	.001
stepVOT	-0.020	0.020	57540	-1.041	.298
PoA	-0.086	0.011	57900	-7.811	<.001
TargetVoice	0.029	0.011	57880	2.711	.007
TargetPoA	0.197	0.011	57880	18.518	<.001
VOT×F_0	-0.048	0.021	33	-2.294	.028
PoA×TargetVoice	0.006	0.011	57880	0.581	.561
PoA×TargetPoA	-0.124	0.011	57880	-11.649	<.001
TargetVoice×TargetPoA	-0.276	0.011	57880	-25.845	<.001
PoA×TargetVoice×TargetPoA	0.040	0.011	57880	3.755	<.001

Note: Results mentioned in the MS are highlighted

Table S9.7. *N1 amplitude as predicted by VOT, F_0, VAS slope, stepVOT, place of articulation (PoA), target voicing, and target PoA for categorical participants only*

	Estimate	*S.E.*	*d.f.*	*t*	*p*
(Intercept)	-1.404	0.319	35	-4.397	<.001
VOT	0.073	0.017	45	4.261	<.001
F_0	0.346	0.071	32	4.911	<.001
stepVOT	0.117	0.019	56540	6.217	<.001
PoA	-0.103	0.010	56550	-9.880	<.001
TargetVoice	0.039	0.010	56550	3.830	<.001
TargetPoA	0.114	0.010	56550	11.191	<.001
VOT×F_0	-0.031	0.028	32	-1.109	.276
PoA×TargetVoice	0.020	0.010	56550	1.922	.055
PoA×TargetPoA	-0.154	0.010	56550	-15.049	<.001
TargetVoice×TargetPoA	-0.178	0.010	56550	-17.430	<.001
PoA×TargetVoice×TargetPoA	-0.032	0.010	56550	-3.111	.002

Note: Results mentioned in the MS are highlighted

References

Kapnoula, E. C., Edwards, J., & McMurray, B. (2021). Gradient activation of speech categories facilitates listeners' recovery from lexical garden paths, but not perception of speech-in-noise. *Journal of Experimental Psychology: Human Perception and Performance*.

McMurray, B. (2017). *Nonlinear Curvefitting for Psycholinguistic (and other) Data*. https://doi.org/none

McMurray, B., Farris-Trimble, A., Seedorff, M., & Rigler, H. (2016). The Effect of Residual Acoustic Hearing and Adaptation to Uncertainty on Speech Perception in Cochlear Implant Users: Evidence From Eye-Tracking. *Ear and Hearing*, *37*(1), e37-51. https://doi.org/10.1097/AUD.0000000000000207

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from lexical garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, *60*(1), 132–158. https://doi.org/10.1016/j.jml.2008.07.002

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, *21*(10), 1532–1540. https://doi.org/10.1177/0956797610384142