The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study

Alan C. L. Yu and Hyunjung Lee[a)]
Department of Linguistics, University of Chicago, Chicago, Illinois 60615

(Received 27 July 2013; revised 17 May 2014; accepted 26 May 2014)

Perceptual compensation for coarticulation (PCCA) refers to listener responses consistent with perceptual reduction of the acoustic effects of the coarticulatory context on a target sound. The robustness of PCCA across individuals and across tasks have not been studied together previously. This study reports the results of two experiments designed to determine the robustness of perceptual compensation for vocalic influence on sibilant perception across tasks and the stability of such compensatory response within an individual. Identification and discrimination data, collected in the laboratory and on Amazon's Mechanical Turk, showed that individuals are moderately stable in their PCCA responses across tasks and the level of stability is consistent across both the lab-based and the internet-based cohorts, although some differences are observed.
 [http://dx.doi.org/10.1121/1.4883380]

PACS number(s): 43.71.Es, 43.71.Sy [BRM] Pages: 382–388

I. INTRODUCTION

Perceptual compensation for coarticulation (PCCA) refers to listener responses consistent with perceptual reduction or perceptual elimination of the acoustic effects of the coarticulatory context on a target sound. A classic demonstration of PCCA is the study of Mann and Repp[1] of vocalic effects on sibilant perception, where they asked listeners to classify sibilants in a series of CV continua where C is a synthesized continuum of /s/ to /ʃ/ and V is either /a/ or /u/. They found that listeners were more likely to report hearing /s/ before /u/ than before /a/, presumably because listeners took into account the lowered noise frequencies of /s/ in a rounded vowel context. Similar findings of perceptual compensation have been observed for other coarticulatory processes, such as vowel-to-vowel dependencies,[2] intrinsic vowel pitches,[3] and vowel nasalization.[4,5]

While the mechanisms underlying PCCA is a matter of much debate,[6–8] most agree that understanding the nature of PCCA has significant ramifications for the understanding of sound change. In particular, many have argued that, when listeners fail to take into account ambiguities, or "noise," in the speech signal, which may stem from articulatory, acoustic, auditory, and perceptual constraints inherent to the vocal tract and the auditory/perceptual apparatus, misperception (i.e., the seed for sound change) may result.[9,10] For example, when the spectral frequency lowering effect of vocalic rounding on /s/, which results in the sibilant having a more /ʃ/-like percept, is not factored out properly by the listeners via the mechanism of PCCA, such listeners might erroneously attribute the lowered frequencies as intentional and reconstruct the sibilant as /ʃ/. This approach makes two important assumptions. To begin with, it assumes that PCCA is the norm and that deviation from this norm is taken to be a matter of perceptual errors. Second, this type of haphazard perceptual errors may accumulate to lead to systematic changes in the language. The time is ripe for revisiting these assumptions in light of recent work that suggests that individuals may differ in the magnitude of PCCA and other context-dependent speech processing effects.[5,11–14]

Individual variability in perception and production has been argued to be important for the understanding of the actuation of sound change,[12,14–16] particularly if the variation in perceptual and production norms across individuals within a population is systematic. Yet, little is known regarding whether the individual differences are stable regardless of task types. To address these questions, we report the results of two experiments, designed to determine the relationship and consistency in PCCA magnitude across tasks (identification and discrimination) within an individual. Perceptual research has long documented a tight relationship between listeners' identification and discrimination functions in speech perception.[17] Little is known about the relationship between identification and discrimination with respect to PCCA, however. To the best of our knowledge, Repp[18] is the only study that looks at individual variability in context-dependent perceptual discrimination. Based on perceptual discrimination data of sibilants presented in various vocalic contexts and in isolation, Repp[18] concludes that there exists two different strategies of listening to fricative-vowel syllables, one auditory, which segregates the noise portion from the vocalic portion, and the other phonetic, where sibilant noise information is more integrated with the vocalic portion. The robustness of this distinction requires further examination, especially given that only two of the ten subjects tested exhibit the auditory mode of listening. The current study aims to expand our understanding on individual variability in speech perception in two important ways. First, we examined directly the relationship between phoneme identification and discrimination in PCCA. Previous studies have primarily observed qualitative connections between identification and discrimination in speech perception. To be best of our knowledge, there has never been explicit correlational

a)Author to whom correspondence should be addressed. Electronic mail: hyunjunglee123@gmail.com

 0001-4966/2014/136(1)/382/7/$30.00

investigation between these two types of perceptual responses. Second, we aim to contribute to the advancement of new methodological technique for investigating individual variability in speech. In addition to traditional data collection methods (i.e., lab-based testing), we also investigated the utility of crowdsourced methodologies for conducting speech perception research over the internet. In addition to data gathered in the laboratory, we also collected perceptual data on Amazon's Mechanical Turk, a Web application that provides instant access to thousands of potential participants for survey-based psycholinguistic experiments.[19,20] Cross-validation studies are important since little is known about the comparability between experimental data collected via the internet and data collected in the laboratory.[20] If the crowdsourcing method for speech perception research proved viable, it stands to be a real boon for investigating the mechanisms underlying individual variation in language and speech, which generally requires the testing of a large sample to achieve statistical power.

II. METHODS

A. Participants

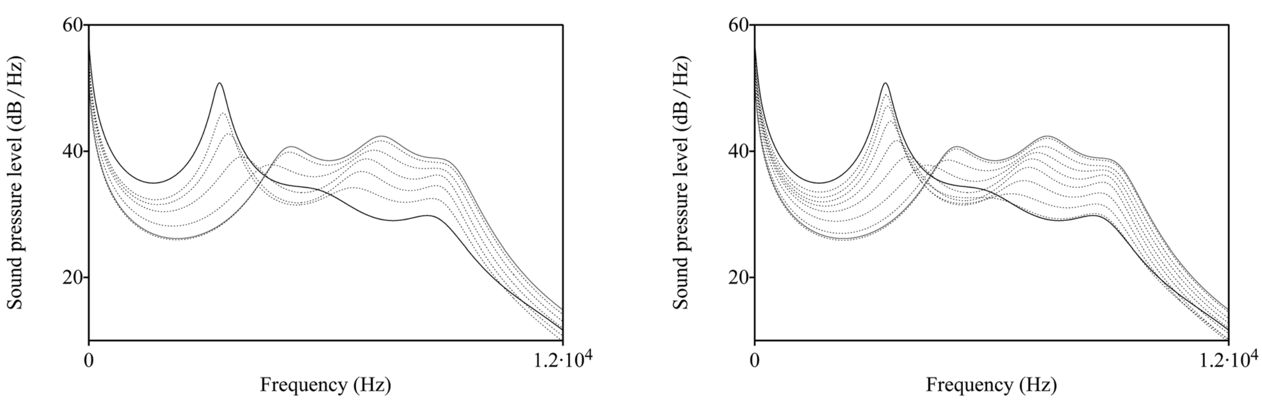
Two cohorts of participants completed two PCCA tasks, a two-alternative forced choice (2AFC) identification task and an auditory AX discrimination task.[21] Seventy-seven adults [48 females, mean age of 23.61 (SD = 7.99)] completed the lab-based study either for course credits or a nominal fee ($5 for 30 min of testing). 187 adults [121 females, mean age of 30 (SD = 10.02)] completed the study on Mechanical Turk for a nominal fee ($2 for approximately 30 min of testing). The participants were all self-reported native speakers of American English without any language and hearing problems, and all of them completed both PCCA tasks.

B. Stimuli

Two sets of /sV – ʃV/ continua were created (V = /a/ or /u/). The sibilant portion of the stimuli came from either a seven-step (the 2AFC task) or a ten-step (the AX task) continuum, synthesized in PRAAT (Ref. 22) using a custom-made script. The script creates a 50-step /s/ ∼ /ʃ/ continuum by digitally mixing in 2% increments various mixtures (a weighted average of the waveforms) of the /s/ and /ʃ/ sounds taken from clear tokens of /sa/ and /ʃa/ produced by a male native speaker of American English. Seven native speakers of English then listened to the fifty resynthesized sibilants and were asked to decide whether the sibilant is /s/ or /ʃ/. The identification results grouped into seven or ten bins according to percentage of /ʃ/-response. Steps 2–6 of the seven-step continuum used in the 2AFC task were selected from bin 2 to bin 6 out of seven bins, while steps 2 to 9 of the ten-step continuum used in the AX task were selected from bin 2 to bin 9 out of ten bins. The original /s/ and /ʃ/ were included as endpoints of the seven-step/ten-step series. The long term average spectra of the sibilants are illustrated in Fig. 1. The center of gravity measures (in Hz) were 6814, 6788, 6519, 5867, 4962, 4177, and 3537 for the seven-step continuum and 6814, 6810, 6677, 6373, 5867, 5226, 4477, 3942, 3655, and 3537 for the ten-step continuum. The sibilants (synthesized and natural) were then cross-spliced with /a/ and /u/ taken from original /da/ and /du/ syllables produced by a male native speaker of American English, which generated two /V_isV_i – V_iʃV_i/ continua for the 2AFC task and two /sV – ʃV/ continua (V = /a/ or /u/) for the AX task. The resulting tokens were then normalized for intensity and pitch. The stimuli may be downloaded at https://db.tt/irY3BaPc (last viewed May 24, 2014).

C. Procedure

The lab-based participants completed the 2AFC and AX tasks over E-Prime in a sound-proof booth. In the 2AFC task, the participants listened to a series of V_iCV_i sequences (C = a synthesized seven-step /s/ – /ʃ/continuum; V = /a/ or /u/) and deciding whether the fricative was /s/ or /ʃ/. Participants were given 3 s to respond before the presentation of the next stimulus. A total of 84 trials



(a) 7-step continuum

(b) 10-step continuum

FIG. 1. Long term average spectra of the (a) seven-step and (b) ten-step /s/ ∼ /ʃ/ continua. The spectrum of the original /s/ is indicated by the solid gray line while the spectrum of the original /ʃ/ is indicated by the solid black line. The spectra of the synthesized sibilants are represented with dotted lines.

(7 continuum steps × 2 vowel contexts × 6 repetitions) were randomly presented. The order of the response options was counter-balanced across participants.

The target stimuli for the AX task were pairs of CV syllables (C ranges from /s/ to /ʃ/ and V = /a/ or /u/). On each trial, a CV pair was presented with one of two interstimulus intervals (ISI): 50 and 750 ms. Participants were instructed to attend to the consonant and indicate whether the two consonants were different using buttons labeled SAME and DIFFERENT (button positions were counter-balanced). Participants were told target consonants would always sound similar and that they should respond SAME only if they hear the targets as identical. On each trial, one target consonant was followed by /u/ and the other by /a/. The target consonants were either identical (catch trials) or differed by three steps along a ten-step series (e.g., step 1 vs step 4, step 2 vs step 5, etc.; discrimination trials). The effect of context was tested by comparing two conditions defined by the arrangement of the targets and the accompanying vowels in each trial. In the "enhanced" condition, target consonant with high center frequency (at /s/-end of the /s/ − /ʃ/ continuum) were followed by the vowel /u/ and target stimuli with low center frequency (at the /ʃ/-end of the same /s/ − /ʃ/ continuum) were followed by the vowel /a/. In the "diminished" condition, the opposite arrangement is used. Based on earlier findings regarding the effects of vocalic contexts on sibilant perception, the discrimination of the target pairs was expected to be more accurate in the "enhanced" condition than in the "diminished" condition. The within-trial order of the CV pairs was counterbalanced to yield 28 unique discrimination trials and 20 unique catch trials. We also included a "control" condition where the target stimuli are embedded in the same vocalic environment. Finally, the natural /da/ and /du/ syllables were paired with original /s/ and /ʃ/ to create four filler pairs with an ISI of 750 ms to enhance the alertness of the participant during the task. All 100 trials ([7 discrimination pairs × 2 conditions (enhanced vs diminished) + 10 catch pairs] × 2 orders (/a/-final syllable first or /u/-final syllable first) × 2 ISIs (50 ms vs 750 ms) + 4 fillers) were randomly presented within a single block and the trial block was repeated four times. Participants were given a short break after two blocks.

For the online tests, participants completed the two tasks on Amazon's Mechanical Turk. The setup of the 2AFC task was identical to that of the lab-based task except that there is no time constraint between the presentation of each stimulus. For the AX discrimination task, the online version was shortened to include only one ISI (750 ms), yielding four blocks of 52 trials ([7 discrimination pairs × 2 conditions (enhanced vs diminished)+10 catch pairs] × 2 orders (/a/-final syllable first or /u/-final syllable first) × 1 ISI (750 ms) + 4 fillers).

III. RESULTS

Subjects' performances were modeled using logistic mixed-effects regression fitted in R, using the *lmer*() function from the LME4 package (version 0.999999-2).[23] The Wald's Z test, which describes how distant a coefficient estimate is from zero in terms of its standard error, was used to test the significance of estimates of the model. The dependent variable is /ʃ/-response (/ʃ/ = 1, s = 0) for the 2AFC task and discrimination accuracy for the AX task (accurate response = 1). To reduce multicollinearity between predictors, continuous variables were centered, categorical variables were sum-coded unless specified otherwise.

TABLE I. Estimates for predictors in a mixed-effects model in the 2AFC identification task. The model formula in LME4 style was /ʃ/ − RESPONSE ∼ TRIAL + STEP × VOWEL × COHORT + (1 + TRIAL + STEP × VOWEL|SUBJECT). Vowel: /a/ = 1, /u/ = −1, and COHORT: lab-based = 1, online = −1.

Predictor	Coefficient β	SE (β)	z	p-value
Intercept	1.549	0.123	12.641	<0.001
TRIAL	0.057	0.038	1.492	= 0.136
STEP	6.378	0.210	30.422	<0.001
VOWEL$_a$	0.508	0.058	8.748	<0.001
COHORT$_{lab}$	−0.271	0.121	−2.233	= 0.026
STEP:VOWEL$_a$	−0.450	0.097	−4.621	<0.001
STEP:COHORT$_{lab}$	−0.660	0.207	−3.185	= 0.001
VOWEL$_a$:COHORT$_{lab}$	0.073	0.058	1.250	= 0.211
STEP:VOWEL$_a$:COHORT$_{lab}$	0.238	0.097	2.439	= 0.015

A. Results of the 2AFC identification task

Following Stewart and Ota,[11] data from five lab-based participants (6%) and 31 (17%) of the internet-based participants were excluded as a result of not reaching at least 50% correct for the two continuum endpoints in the 2AFC task.

The regression model contains four predictors: TRIAL indexed the order in which the particular utterance was produced, STEP indexed a stimulus' fricative location on the /s/ − /ʃ/ continuum, VOWEL the nature of the following vowel (/a/ = 1, /u/ = −1), and COHORT the test population (lab-based = 1, online = −1). The model includes a by-SUBJECT random intercept to allow for subject-specific variation in /ʃ/ response rate, as well as all possible by-SUBJECT random slopes to allow for subject-specific variation for the fixed effects.

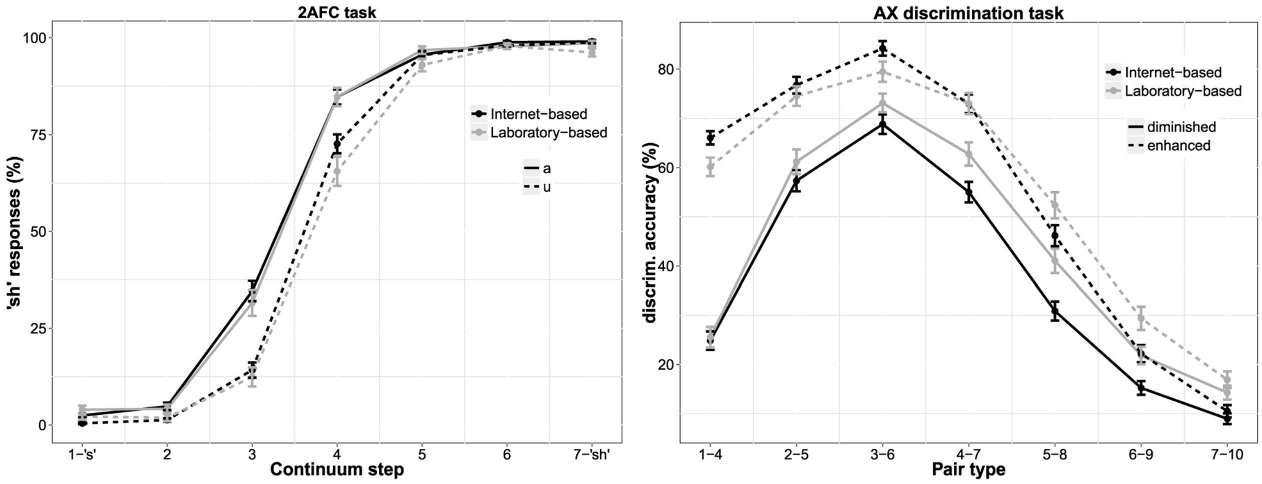
Table I summarizes the parameter estimate for each of the fixed effects in the model, as well as the estimate β of its standard error SE(β), the associated Wald's Z-score, and the significance level. Main effects of STEP and VOWEL were significant as expected: the rate of /ʃ/ responses increased with increasing STEP ($\beta = 6.38$, $p < 0.001$); /ʃ/ response was lowest when the following vowel is /u/ ($\beta = 0.51$, $p < 0.001$). A significant interaction between STEP and VOWEL was found ($\beta = -0.45$, $p < 0.001$), suggesting that the influence of vowel on the /ʃ/ response varies across continuum step. As illustrated in Fig. 2(a), the vocalic influence is strongest in the middle of the continuum than at the endpoints of the continuum. To examine potential differences in perceptual behavior across cohorts, the interaction of COHORT with the significant fixed factors were tested. There is a significant main effect of COHORT ($\beta = -0.27$, $p < 0.05$), suggesting that the baseline /ʃ/-response rate is lower among the lab cohort than the online cohort. However, there is not a significant interaction between COHORT and

Vowel ($\beta = 0.07$, $p = 0.21$), suggesting that the two cohorts do not differ in their PCCA responses, the primary effect of interests here. Cohort does significantly interact with Step ($\beta = -0.66$, $p < 0.01$), although this interaction is further mediated by the vocalic context, as indexed by a three-way interaction between Cohort, Step, and Vowel ($\beta = 0.24$, $p < 0.05$). As illustrated in Fig. 2(a), the slope of the identification function in the /u/-context is shallower for the lab cohort than the online cohort. The reason behind this cohort difference is not clear. A shallower identification function suggests more ambiguities between /s/ and /ʃ/ in the /u/ context among the lab cohort. It should be noted that /u/ varies greatly in rounding across different regional varieties of English in the United States. In particular, it has been suggested that fronted /u/ might have come to cover "90% of the North American continent" (Labov,[24] p. 27). While further investigation is needed, the steeper identification function among the online cohort might reflect less variability (i.e., less rounding influence) in the realization of /s/ and /ʃ/ before /u/ across the population a whole.

B. Results of the AX discrimination task

For the AX task, data from ten lab-based participants (13%) and eight internet-based participants (4%) were excluded from further analysis since they did not reach at least 50% for the catch trials. The regression model contains four predictors: Trial indexed the order in which the stimuli were presented, discrimination Pair indexed pairs of sibilant-vowel combination along the ten-step /s/ – /ʃ/ continuum, Condition the perceptually-enhancing or diminishing arrangement of vowels (diminished = 1, enhanced = −1), and Cohort the test population (lab-based = 1, online = −1). To reduce model complexity, the Pair variable was reduced to three levels, with pairs 4∼7, 5∼8, and 6∼9 as Level 1 (i.e., pairs from the most ambiguous region of the /s/ ∼ /ʃ/ continuum), pairs 1∼4 and 2∼6 as Level 2, and pairs 6∼9 and 7∼10 as Level 3. The Pair variable was further contrast-coded such that Contrast 1 compares Level 1 with the average of the other two levels, while Contrast 2 compares Level 2 with Level 3 (i.e., pairs closest to the /s/ and /ʃ/-end of the continuum, respectively). The model includes a by-Subject random intercept to allow for subject-specific variation in accuracy rate, as well as by-Subject random slopes for Trial, Pair, and Condition, as well as the interaction of Pair, and Condition, to allow for subject-specific variation for the fixed effects. Interstimulus interval was dropped in this analysis as a likelihood ratio test comparing between a model with and without the ISI interval factor was not significant [$\chi^2(1) = 0.04$, $p = 0.85$].

Table II summarizes the parameter estimate for each of the fixed effects in the model, as well as the estimate β of its standard error SE(β), the associated Wald's Z-score, and the significance level. Main effects of Pair and Condition were significant as expected. As illustrated in Fig. 2(b), discrimination accuracy is highest in the middle region of the /s/ ∼ /ʃ/ continuum ($\beta = 1.46$, $p < 0.001$). Discrimination accuracy is also higher at the /s/-end of the continuum than the /ʃ/-end of the continuum ($\beta = 2.31$, $p < 0.001$), suggesting that listeners are more sensitive to changes in the /s/-end of the continuum than the /ʃ/-end. Crucially, there is a main effect of Condition ($\beta = -0.42$, $p < 0.001$), suggesting that listeners are more accurate at sibilant discrimination in perceptually enhancing contexts than in diminishing ones. A significant interaction between Pair and Condition suggests that the enhancement effect differs across stimulus pairs. The enhancement effect is stronger in the middle of the continuum, where the sibilants are maximally ambiguous, relative to the average accuracy rates at the two ends of the continuum ($\beta = 0.09$, $p < 0.01$); it is also stronger toward the /s/-end of the continuum than the /ʃ/-end of the continuum ($\beta = -0.51$, $p < 0.001$). This difference in enhancement might reflect a stronger coarticulatory influence of vowel on /s/ than on /ʃ/ in production. The main effect of



(a) 2AFC Task (b) AX Task

FIG. 2. (a) Mean percentage of /ʃ/ response in /u/ (dashed line) and /a/ (solid line) contexts by the lab-based (dark color) and internet-based (light color) cohorts. (b) Mean discrimination accuracy in the "enhanced" (dashed line) and "diminished" (solid line) conditions by the lab-based (dark color) and internet-based (light color) cohorts.

TABLE II. Estimates for predictors in a mixed-effects model in the AX discrimination task. The model formula in LME4 style was RESPONSE ACCURACY ~ TRIAL + PAIR × CONDITION + PAIR × COHORT + (1 + TRIAL + PAIR × CONDITION|SUBJECT). CONDITION: diminished = 1, enhanced = −1, COHORT: lab-based = 1, online = −1, PAIR$_{contrast1}$: pairs 4~7, 5~8, and 6~9 vs the average of pairs 1~4 and 2~6 and pairs 6~9 and 7~10; PAIR$_{contrast2}$: pairs 1~4 and 2~6 and pairs 6~9 and 7~10.

Predictor	Coefficient β	SE (β)	z	p-value
Intercept	−0.272	0.036	−7.517	<0.001
TRIAL	0.116	0.021	5.563	<0.001
PAIR$_{contrast1}$	1.458	0.061	23.814	<0.001
PAIR$_{contrast2}$	2.308	0.122	18.850	<0.001
CONDITION$_{diminished}$	−0.418	0.023	−18.084	<0.001
COHORT$_{lab}$	0.206	0.036	5.716	<0.001
PAIR$_{contrast1}$: CONDITION$_{diminished}$	0.089	0.027	3.231	= 0.001
PAIR$_{contrast2}$: CONDITION$_{diminished}$	−0.512	0.041	−12.562	<0.001
PAIR$_{contrast1}$:COHORT$_{lab}$	−0.008	0.055	−0.151	= 0.880
PAIR$_{contrast2}$:COHORT$_{lab}$	−0.319	0.122	−2.617	= 0.009

TRIAL suggests that discrimination accuracy increased as the experiment progressed ($\beta = 0.12$, $p < 0.001$). Of particular interest here is the effect of COHORT. While we see a significant main effect of COHORT ($\beta = 0.21$, $p < 0.001$), suggesting that the lab-based cohort has a higher accuracy rate than the online cohort, a likelihood ratio test comparing between a model with and without the CONDITION × COHORT interaction suggests that the enhancement effect does not differ across the two cohorts [$\chi^2(1) = 2.27$, $p = 0.13$]. The addition of an interaction between PAIR and COHORT was significant [$\chi^2(2) = 6.66$, $p < 0.05$], showing in particular that the difference in accuracy between the /s/- and /ʃ/-end of the continuum was less pronounced in the lab-based cohort than in the online one (CONTRAST 1 × COHORT: $\beta = -0.01$, $p = 0.88$; CONTRAST 2 × COHORT: $\beta = -0.32$, $p < 0.01$).

C. Comparison across tasks

To determine whether PCCA is stable across individuals and across tasks, we examined the relationship between a participant's PCCA responses in the 2AFC task and his/her performance in the AX discrimination task. To measure the effects of PCCA in the 2AFC task, we subtracted the mean percentage of /ʃ/ response at each step along the /a/-continuum from its counterpart along the /u/-continuum and calculated the mean of the seven measures for each participant. To measure the effects of PCCA in the AX task, we subtracted the mean discrimination accuracy rate for each stimulus pair in the "enhanced" condition from its counterpart in the "diminished" condition and calculated the mean of the seven measures for each participant.

As shown in Fig. 3, the magnitude of PCCA in the 2AFC task is correlated with that of the AX task regardless of cohort population [r (205) = 0.346, $p < 0.001$]. There is a significant positive correlation between the degree of PCCA in the lab-based 2AFC and the AX tasks [r (54) = 0.28, $p = 0.04$] as well as an even stronger positive correlation

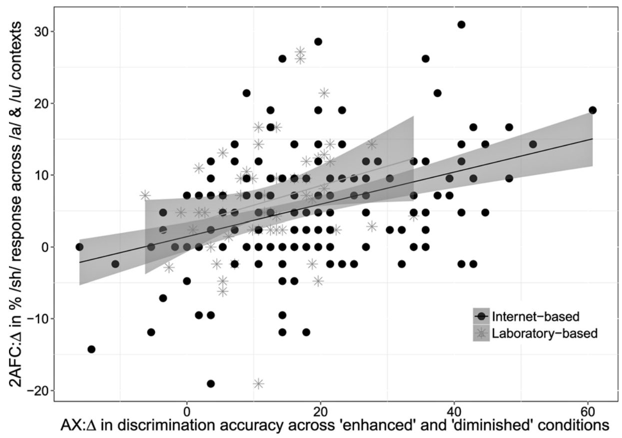


FIG. 3. Correlation between perceptual compensation responses across the 2AFC and AX tasks; lab-based: r (54) = 0.28, $p = 0.04$; internet-based: $r(149) = 0.39$, $p < 0.001$; both cohorts combined: r (205) = 0.346, $p < 0.001$.

between the degree of PCCA in the internet-based 2AFC and AX tasks [r (149) = 0.39, $p < 0.001$]. A two-tailed Fisher's r to z transformation shows that there is not a significant difference in the lab-based and internet-based correlations ($z = -0.66$, $p = 0.36$). These results suggest that participants vary greatly in the magnitude of their PCCA responses, not only in individual tasks, but also across tasks. This individual variability is not random, however. While the strength of the across-task correlation is moderate, our findings suggest that listeners do show some level of consistency in their PCCA responses, as evidenced by the fact that participants who exhibit strong PCCA responses in one task are likely to exhibit similarly strong PCCA responses in the other.

IV. DISCUSSION AND CONCLUSION

This study examined PCCA across two behavioral tasks, showing that the participants as a group, regardless whether they participated in the lab or online, exhibit clear PCCA effects. The picture is considerably more complicated when looking at PCCA responses at the individual level. While listeners appear to be somewhat stable in their PCCA responses across identification and discrimination tasks, as seen in Sec. III C, there exists nonetheless plenty of variation across individuals in their PCCA responses within and across tasks. To begin with, individuals do not compensate for coarticulatory effects in the same magnitude. Some are robustly compensating while others weakly if at all. This echoes the findings of Repp[19] where listeners seem to exhibit two modes of listening. Some appear to pay closer attention to the auditory properties of speech and not taking the vocalic contexts into account, while others appear to be engaging in phonetic listening, thus allowing the listener's implicit knowledge of articulation and coarticulation and/or their acoustic consequences to affect perceptual decisions. However, rather than a clear dichotomy that was suggested by this two-listening-mode hypothesis, our findings suggest that listeners seem to vary along a continuum. Further investigation is needed to ascertain the perceptual mechanisms that underlie this listening-mode continuum (cf. Yu[12]). Besides across individual variability within task, individuals also seem to exhibit fluctuation in PCCA across tasks. That is, while the correlation between PCCA responses in identification and discrimination tasks are significantly correlated, the strength of the correlation is moderate at best. This suggests that other factors, such as fatigue and general fluctuation in attentional focus, might be modulating the level of PCCA responses in each task. Further investigation is needed to ascertain what factors might modulate PCCA.

While further investigation is needed to determine whether the type of stable PCCA observed in this study is specific to vocalic influence on sibilant or whether this is a general phenomenon that transcends all coarticulatory effects, to the extent that the present findings are robust, they hold serious implications for the listener-misperception view of sound change. To begin with, our findings suggest that differences in perceptual norms need not come from the accumulation of perceptual errors. PCCA variation exists *in rerum natura*. Listeners who compensate weakly may not reconstruct the same sound category under coarticulatory influence as would those who compensate more strongly; there are many shades of compensatory response strengths, to be sure. In particular, individuals with weaker PCCA tendencies are predicted to be the type of listeners that would innovate new variants since context-dependent variants are perceived as if they are inherent (e.g., a word that sounds halfway between [su]or [ʃu] might be analyzed by listeners who are weakly compensating as /ʃu/ underlyingly and as /su/ by someone who more strongly compensated).

The above scenario might also explain why sound change is systematic, both in the sense of within individual idiolects and also within a population as a whole. Traditional listener-misperception view of sound change assumes that sound change is lexically abrupt.[9] A perceptual error affects the lexical item in question and no others. Systematic changes of the sound system are assumed to be the results of perceptual error accumulation over time and the diffusion of such changes across the lexicon. Our findings suggest a difference source of systematicity. Within the individual, listeners may show, moderate as it may be, some level of consistency in the magnitude of perceptual compensation for context-induced variation. This type of internal consistency within an idiolect exists presumably because it is constructed out of the same perceptual apparatus. Thus from the point of view of sound change propagation, our findings suggest that, not only sound change need not rely on the adoption of haphazard perceptual errors by others, there might also be regularities at the population level as well. That is, while cross-individual variability provides the linguistic resources needed for the construction of linguistic styles, it is crucially the individuals who produce those linguistic variants to which socio-indexical meaning will anchor. If a subsegment of a population consistently do not compensate for context-induced variation in speech, their perceptual and production norms have the potentials to be treated by the early adopters[25] as population norms and subsequently propagate the innovative variants to the rest of the speech community. Future sociolinguistic investigations might examine whether innovators are more likely to be noncompensators.

Finally, our findings also contribute to the growing literature on crowdsourced perceptual research. Given the need for statistical power in perceptual research that focuses on individual variability, it is paramount to determine whether crowdsourcing is a viable method for conducting speech perception and production research. While there exists intriguing differences, such as the effects of continuum step across population, our findings suggest that perceptual data obtained online yields results that are consistent with data obtained in traditional laboratory settings, at least with respect to the primary experimental manipulation of interests. This cross-validation study thus lends further support for the utility of internet-based speech perception studies.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation Grant No. BCS-0949754.

[1] V. A. Mann and B. H. Repp, "Influence of vocalic context on perception of the [sh]-[s] distinction," Percept. Psychophys. **28**, 213–228 (1980).

[2] P. S. Beddor, J. Harnsberger, and S. Lindemann, "Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates," J. Phonetics **30**, 591–627 (2002).

[3] J.-M. Hombert, '"Development of tones from vowel height?," J. Phonetics **5**, 9–16 (1977).

[4] P. Beddor and R. A. Krakow, "Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation," J. Acoust. Soc. Am. **106**(5), 2868–2887 (1999).

[5] P. S. Beddor, "A coarticulatory path to sound change," Language **85**(4), 785–832 (2009).

[6] J. Elman and J. McClelland, "Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes," J. Memory Lang. **27**, 143–165 (1988).

[7] C. Fowler, "Compensation for coarticulation reflects gesture perception, not spectral contrast," Percept. Psychophys. **68**(2), 161–177 (2006).

[8] A. Lotto and K. Kluender, "General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification," Percept. Psychophys. **60**(4), 602–619 (1998).

[9] J. Ohala, "The phonetics of sound change," in *Historical Linguistics: Problems and Perspectives*, edited by C. Jones (Longman Academic, London, 1993), Chap. 9, pp. 237–278.

[10] J. Blevins, *Evolutionary Phonology: The Emergence of Sound Patterns* (Cambridge University Press, Cambridge, 2004).

[11] M. E. Stewart and M. Ota, "Lexical effects on speech perception in individuals with 'autistic' traits," Cognition **109**, 157–162 (2008).

[12] A. C. L. Yu, "Perceptual compensation is correlated with individuals' 'autistic' traits: Implications for models of sound change," PLoS One **5**(8), e11950 (2010).

[13] A. C. L. Yu, J. Grove, M. Martinović, and M. Sonderegger, *Effects of Working Memory Capacity and "Autistic" Traits on Phonotactic Effects in Speech Perception*, edited by E. Zee (International Congress of the Phonetic Sciences, Hong Kong, 2011), pp. 2236–2239.

[14] A. C. L. Yu, *Individual Differences in Socio-cognitive Processing and the Actuation of Sound Change*, edited by A. C. L. Yu (Oxford University Press, Oxford, UK, 2013), Chap. 10, pp. 201–227.

[15] A. Baker, D. Archangeli, and J. Mielke, "Variability in American English s-retraction suggests a solution to the actuation problem," Lang. Var. Change **23**(3), 347–374 (2011).

[16] A. Garrett and K. Johnson, *Phonetic Biases in Sound Change*, edited by A. C. L. Yu (Oxford University Press, Oxford, 2013), Chap. 3, pp. 51–97.

[17] A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **54**, 358–368 (1957).

[18] B. H. Repp, "Two strategies in fricative discrimination," Percept. Psychophys. **30**(3), 217–227 (1981).

[19] E. Gibson, S. Piantadosi, and K. Fedorenko, "Using Mechanical Turk to obtain and analyze English acceptability judgments," Lang. Ling. Compass **5**(8), 509–524 (2011).

[20]J. Sprouse, "A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory," Behav. Res. Methods **43**(1), 155–167 (2011).
[21]J. D. W. Stephens and L. L. Holt, "Preceding phonetic context affects perception of nonspeech," J. Acoust. Soc. Am. **114**(6), 3036–3039 (2003).
[22]P. Boersma and D. Weenink, PRAAT [computer program] (2011).
[23]D. Bates, M. Maechler, and B. Bolker, LME4, R package version 0.999375-38 [computer program] (2011).
[24]William Labov, in *Triggering Events*, edited by S. Fitzmaurice and D. Minkova (Mouton de Grutyer, Berlin, 2008), pp. 11–54.
[25]J. Milroy and L. Milroy, "Linguistic change, social network and speaker innovation," J. Ling. **21**(2), 339–384 (1985).