Picture Naming or Word Reading: Does the Modality Affect Speech Motor Adaptation and Its Transfer?

Tiphaine Caudrelier, Pascal Perrier, Jean-Luc Schwartz, Amélie Rochet-Capellan

Interspeech 2018
2-6 September 2018, Hyderabad

Picture naming or word reading: does the modality affect speech motor adaptation and its transfer?

Tiphaine Caudrelier[1], *Pascal Perrier*[1], *Jean-Luc-Schwartz*[1], *Amélie Rochet-Capellan*[1]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-lab, 38000 Grenoble, France
* Institute of Engineering Univ. Grenoble Alpes
Tiphaine.caudrelier@gipsa-lab.fr, Amelie.rochet-capellan@gipsa-lab.fr

Abstract

Auditory-motor adaptation and transfer paradigms are increasingly used to explore speech motor control as well as phonological representations underlying speech production. Auditory-motor adaptation is generally assumed to occur at the sensory-motor level. However, few studies suggested that linguistic or contextual factors such as the modality of presentation of stimuli influences adaptation.

The present study investigates the influence of the modality of stimuli presentation (written word vs. a picture representing the same word) on auditory-motor adaptation and transfer. In this speech production experiment, speakers' auditory feedback was altered online, inducing adaptation. We contrasted the magnitude of adaptation in these two different modalities and we assessed transfer from /pe/ to the French word /epe/ in the same vs. different modality of presentation, using a mixed 2*2 subject design.

The magnitude of adaptation was not different between modalities. This observation contrasts with recent findings showing an effect of the modality (a written word vs. a go signal) on adaptation. Moreover, transfer did occur from one modality to the other, and transfer pattern depended on the modality of transfer stimuli. Overall, the results suggest that picture naming and word reading rely on sensory-motor representations that may be linked to contextual (or surface) characteristics.

Index Terms: auditory perturbations, adaptation, transfer, modality, picture naming, phonological units

1. Introduction

Adaptation to formants perturbation is increasingly used to investigate speech motor control in adults and children, including typical and atypical speakers [1]–[3]. This paradigm consists in applying a real-time systematic perturbation to one or several formants in speakers' auditory feedback. With training, the perturbation induces a change in speakers' pronunciation, called adaptation. This adaptation may or may not transfer to untrained speech material [4]. In most auditory-motor adaptation experiments, utterances were prompted by written words, but studies involving children or illiterate adults rather used pictures [5]. It is however unclear if the modality of presentation (written word vs. picture) affects auditory-motor adaptation and its transfer to other utterances. In other words, is the access to phonological representations and their associated sensory-motor characterizations mediated in the same way according to whether the stimuli are written-words or pictures?

Effects of non-linguistic properties of stimuli (or surface characteristics effects) have been found in research on memory and language [6]. For example, the typography of stimuli (hand-written in uppercase vs. typed in lowercase) was shown to influence priming effect in a word-fragment completion task [7]. In speech production, Ferrand et al. [8] found that written word primes facilitated both word and picture naming. Moreover the authors showed that priming effects in picture naming relied mostly on phonological representations of whole words, while priming in word reading was primarily sensitive to orthographic overlap between the prime and the target [8]. Hence, reading vs. picture naming may activate different processes and representations [9].

The effect of modality observed in priming studies questions the use of pictures vs. written words as neutral factors in auditory-motor adaptation studies, and in particular studies using formant perturbations and analyzing transfer of adaptation. Indeed, the priming paradigm and the auditory-motor adaptation transfer paradigm have in common that they aim at revealing common speech representations between different stimuli, words or situations. These words are the "prime" and the "target" in the former, whereas the latter reveals common representations between a training word and a testing word. In auditory-motor adaptation studies, speakers produce one or several training word(s) while one or several formants are shifted in real time in their auditory feedback [4], [10], [11]. Over repetitions, most speakers adapt to the perturbation: they learn to partially compensate for it by changing their production in the opposite direction of the perturbation (e.g., in the production of "head", when F1 and F2 are shifted toward "had", production is adapted toward "hid"). The pronunciation change persists at least transiently when the perturbation is removed either for the pronunciation of the training material (after-effect) or for the pronunciation of other speech material (transfer effect). The study of transfer profiles was introduced as a way to question the nature of speech production representations [4], [11]–[15]. More specifically, contrasting transfer at the level of different linguistic units (utterances, syllables, and phonemes) was used to reveal speech representations interfacing with speech articulation [15].

Adaptation to formants perturbation is assumed to change auditory-motor representations. Yet several studies have shown an effect of linguistic factors on this adaptation. In particular, the magnitude of adaptation is influenced by the lexical status of the produced word and heard word, with larger adaptation for pseudo-words shifted to real-words than for real-words shifted to pseudo-words [16]. The frequency of the word which speakers are expected to pronounce to compensate for the perturbation also influences adaptation, with lower amplitude of adaptation if the articulatory target is more frequent [17]. Moreover, in a recent study, Alsius et al. [18] contrasted adaptation occurring when the prompt to produce the training word "head" was a linguistic stimulus (a written word or a

10.21437/Interspeech.2018-1760

spoken word) vs. a go signal (a cross or a tune). Results suggested that linguistic prompts induced more adaptation than non-linguistic ones, regardless of the sensory modality, which was interpreted as evidence that linguistic prompts tend to define more precise speech goals than non-linguistic ones.

All together, these various studies suggest that representations involved in speech articulation might be related both to linguistic and surface factors. To further investigate the link between these different levels, the present study contrasts the magnitude of adaptation to formants perturbation on the French pseudo-word /pe/ in two different modalities of presentation: a written word (W) vs. a picture (P) of a fictional character. We then evaluated the interaction between linguistic and surface effects by testing transfer of adaptation to the French word /epe/ (meaning "sword") and after-effects in the same modality than training (word-to-word (WW) and picture-to-picture (PP)) vs. in different modalities (word-to-picture (WP) and picture-to-word (PW)).

2. Method

2.1. Predictions

Our main hypothesis is that access to orthographic representations interferes with the specification of the phonological content of stimuli. Thus it would predict (1) different adaptation when speakers are trained with a written word (W) than with a picture (P) and (2) different transfer depending on the modality of presentation of the transfer stimulus. In addition, the role of surface characteristics would predict (3) more transfer and after-effect in the same modality as training (WW and PP conditions) than in the other modality (WP and PW conditions).

2.2. Participants, task and auditory feedback perturbation

Twenty-six native speakers of Metropolitan French ranging in age from 18 to 31 years old (mean=23, 11 women), were tested in a soundproof room. All reported no history of speech, language or hearing disorders. They were naïve to the purpose of the experiment.

Talkers were prompted to read words or name pictures displayed on a monitor, while wearing headphones with a built-in microphone (Sennheiser HME 26-II-600). They were instructed to speak aloud in a natural way, without shouting or whispering. They heard their own voice at about 70 dB mixed with speech shaped noise at 50 dB in the headphones.

During the training phase (see experimental procedure described below), the first two formants were shifted online using the Audapter system [19]. F1 was increased by 27% and F2 was decreased by 10%, making the vowel /e/ closer to /ɛ/ in speakers' auditory feedback. The perturbation triggered a non-disruptive 14ms delay.

2.3. Stimuli, experimental design and procedure

Stimuli were either written words or pictures (see Fig.1). As the training word "pé" (/pe/) is not a real French word, a fictional character was used to represent the pseudo-word /pe/ (participants were told that the character was named /pe/). /epe/ was elicited by a drawing of a sword ("épée" means sword in French).

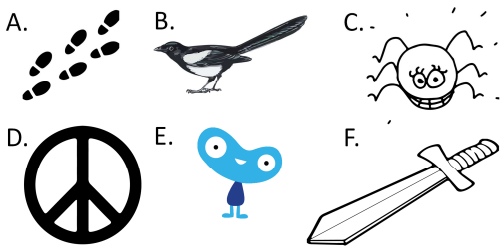
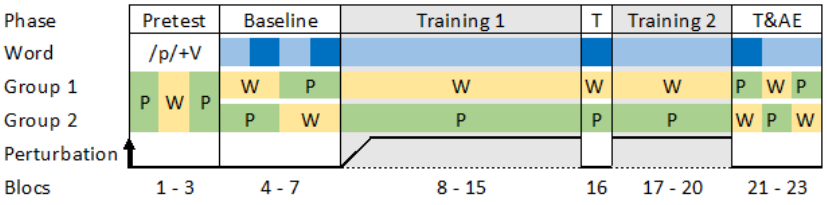


Figure 1: *Pictures A, B, C, D, E, F prompted /pa/ (footprint), /pi/ (magpie), /pu/ (louse), /pɛ/* (peace), the training word /pe/ (a fictional character called /pe/ in the experiment), and the test word /epe/ (sword), respectively. *A,B,C and D were used during the pretest only.*

The experimental design and procedure are illustrated in Figure 2. The experiment consisted of 23 blocks of 20 trials, divided into 6 phases. Between each block, participants could rest and they had to press a key to continue the task. In the pretest, to explore their vocalic triangle, speakers were prompted to pronounce a set of syllables of the form /p/+Vowel, consisting in four real words /pa/, /pi/, /pu/, and /p ɛ/ and one pseudo-word (/pe/).. The /p/+V syllables were prompted alternately by corresponding written words or pictures (see Figure 1). In the baseline (blocks 4 to 7), speakers pronounced 40 times the training utterance (/pe/) and the testing utterance (/epe/) without any perturbation. Half of the utterances were prompted by pictures, and the other half by written words (presented in successive blocks, as specified in Figure 2). In the training phase (blocks 8 to 15), both groups pronounced the training utterance /pe/ while the perturbation was applied to their auditory feedback. The stimuli were the written word "pé" in group 1 and the picture associated with /pe/ in group 2. The first block after training (block 16) tested the transfer to /epe/ in the same modality (word in group 1 and picture in group 2, which corresponds to conditions WW and PP respectively). As the transfer was assessed with no perturbation, adaptation could progressively decay. Therefore, speakers underwent a second training phase (block 17 to 20) to adapt again to the perturbation. Transfer to /epe/ was then assessed in the other

A.



B.

	Transfer		After-effect	
Phase	1	2	1	2
Group 1	WW	WP	WW	WP
Group 2	PP	PW	PP	PW

Figure 2: A. Experimental procedure schematic describing stimuli and perturbation by group over the course of the experiment. B. Experimental conditions by group and by experimental phase

modality (picture in group 1 and word in group 2; WP and PW respectively). Finally, after-effect on /pe/ was evaluated first in the same modality (word in group 1 and picture in group 2, WW and PP respectively) and then in the other modality (picture in group 1 and word in group 2, WP and PW respectively).

2.4. Data analyses

Formants F1 and F2 were evaluated in a 30ms window in the stable part of each vowel. The difference F2-F1 was computed in Hz to assess the magnitude of adaptation, transfer and after-effect as it accounts for changes in both formants. In each speaker, F2-F1 was expressed as a percentage of change compared to the baseline of the same syllable, associated with the exact same stimulus in the same modality of presentation.

Adaptation was evaluated in the last block of each training phase. For each subject, a t-test determined whether the subject significantly adapted to the perturbation or not.

To assess the magnitude of adaptation, transfer and after-effect, we built several Linear Mixed Models (LMM) using the lmer function from the lme4 package of R software. A backward deletion approach [20] enabled to select the best linear mixed model based on likelihood-ratio test. In all LMMs, the change in (F2-F1) was the variable to explain, and the random factor was the speaker. Fixed factors considered were (a) experimental phase and modality of training in the adaptation model; (b) modality of training, modality of transfer and syllable in the transfer model; (c) modality of training and block in the after-effect model. Post-hoc tests were adjusted for multiple comparisons, using the glht function from the multcomp package of R [21].

3. Results

As a result of the auditory feedback perturbation, 11 subjects out of 13 significantly adapted to the perturbation in each group by increasing the difference between their first two formants (F2-F1) at the end of the first training phase in comparison with the baseline. Non-adapted speakers were included in the analysis of adaptation in Section 3.1 (as in previous work on adaptation [18]) but then excluded from the analyses of transfer and after-effect in Sections 3.2 & 3.3, since these phenomena are by construction based on adaptation [15]. In addition, two subjects (one per group) were excluded from all results because they did not follow the instructions properly (one produced the wrong word in the baseline, and the other missed most trials because he was falling asleep).

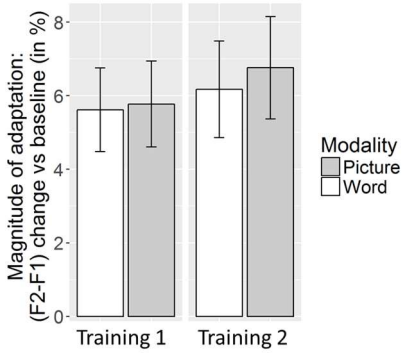


Figure 3: *Magnitude of adaptation by modality of training (written words vs. pictures) and by training phase. Error bars represent standard errors.*

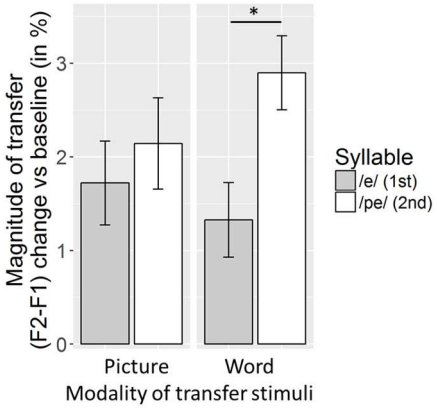


Figure 4: *Magnitude of transfer by modality and by syllable of /epe/. Modalities of adaptation are combined. Error bars represent standard errors. * $p<0.05$*

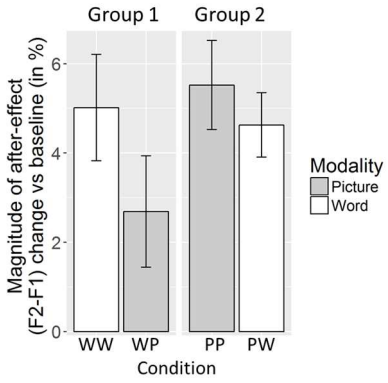


Figure 5: *Magnitude of after-effect averaged by condition. Error bars represent standard errors.*

3.1. Adaptation

Figure 3 displays the magnitude of adaptation in both groups and both training phases. The average adaptation was 6.1%±1.23, with an increasing trend between the first and second training (+0.77%). The best LMM fitting data did not depend on experimental phase although its effect was almost significant (Chisq(1)=3.43; p=0.064). The modality of training (word vs. picture) had no significant effect (Chisq(1)=0.05; p=0.8).

When excluding non-adapted speakers, the increase between training 1 and training 2 stayed not significant (Chisq(1)=3.59; p=0.058).

3.2. Transfer

We analyzed transfer magnitude for the word /epe/, by syllable, by modality of adaptation and by modality of transfer. Some transfer seemed to occur in all conditions. The best LMM fitting transfer data showed that modality of adaptation played no significant role while transfer depended on the interaction between syllable and modality of transfer (see Fig. 4). Post hoc comparisons revealed that transfer was significantly lower in

syllable /e/ than in /pe/ (i.e. first vs. second syllable) when /epe/ was prompted by the word "épée" (i.e. in conditions PW or WW; z=-2.824; p=0.032) but not when prompted by the picture (i.e. in conditions PP or WP: z=-0.812; p=0.917). Post hoc tests also confirmed that transfer magnitude was significantly above 0 in all conditions of adaptation and transfer (p<0.001).

3.3. After-effect

After the second transfer block, when perturbation is stopped, after-effect on the training utterance /pe/ is assessed on the same modality as training and then on the other modality (see Fig. 5). In both groups, after-effect seems to decay between first and second block of after-effect. The best LMM had block as the only factor (Chisq(1)=6.03; p=0.014). Hence globally, WW and PP have higher after-effects than WP and PW.

4. Discussion

In the present study, we assessed the effect of the modality of presentation of stimuli (written words vs. pictures) on sensory-motor adaptation to auditory perturbations. More specifically, we introduced the change in modality between adaptation and its transfer and after-effect, as a way to question the role of "surface effects" [6] in auditory-motor adaptation. As in our previous work [15], we argued that the study of transfer might highlight the nature of the link between phonological content and sensory-motor representations in speech production.

The first question concerned the possible interference of orthographic factors in the phonological specification when training stimuli were written words as opposed to pictures. Results show significant and perceivable [22] adaptation, but no significant difference in the adaptation between modalities. This observation does not corroborate recent findings suggesting that written or spoken words induce more auditory-motor adaptation than non-linguistic "go" signals (such as a visual cross or a tune) [18].

The absence of significant results could have several explanations. It could be that the number of participants in the present study was not large enough to enable significant results to be found (Alsius et al's study [18] included 64 participants vs. 26 participants in the present study). However, there was absolutely no trend related to modalities in our adaptation data, so this interpretation seems unlikely to us. Alternatively, the absence of effect of modality on adaptation could be explained by the nature of the stimulus. Participants were told that the fictional character associated with the training utterance was called /pe/, whereas in [18] the visual go signal was a cross, that was not related to the word. If a name is explicitly given to the picture (or if the picture explicitly refers to a word), it may lead to a better specification of speech goals than in the case of a cross, and then lead to more adaptation. The influence of the nature of the link between the picture and its associated utterance could be investigated in further research.

Transfer was then assessed from /pe/ to the real word /epe/. We tested whether the orthographic partial overlap between the written words ("pé" and "épée") could interfere with transfer (prediction 2) and investigated whether surface characteristics of stimuli were somehow linked to speech motor representations (in relation to prediction 3). The results show that interference between transfer and modality occurs only for transfer modality, but not for adaptation modality, showing that the effect of the syllable (/e/ vs. /pe/) depends on the modality of the test stimuli, regardless of the modality of the training stimuli. As a matter of fact, the difference in transfer magnitude between /e/ and /pe/ in /epe/ is higher when the test stimuli are words (i.e. in WW and PW) than when they are pictures (in WP and PP), and this is not dependent on the modality used during adaptation. The transfer pattern observed with word stimuli is consistent with previous findings on disyllabic words [15]. Higher transfer was observed at a syllabic level (that is from /pe/ alone to /pe/ in /epe/) than at a phoneme level (that is from /pe/ alone to /e/ in /epe/). The interaction between test modalities and syllables may be interpreted in terms of surface characteristics effects [6]. At a phonological level, the second syllable is the same as the training utterance. The orthographic representation of this syllable may activate more intensely phonological representations, modified during training, than in the situation of picture naming. Furthermore, masked priming studies have shown that picture naming relied mostly on whole word phonological representations [8]. If pictures activated only holistic word representations [9], then transfer should not happen nor at a syllabic level neither at a phonemic one – which was not what we observed. The effect of the modality on the level at which transfer occurs the most (phoneme vs syllable vs word) could be investigated using a broader set of training and test words and/or pictures.

Finally, the possibility of a link between surface characteristics of stimuli and speech motor representations was tested in the after-effect. We actually found that there was significantly more after-effect in WW and PP than in WP and PW conditions. The results hence tend to suggest that after-effects are stronger when adaptation and test are made with the same modality. However, this interpretation should be nuanced, since this effect can be a simple effect of the block. Indeed, the after-effect gradually decays over repetitions since there is no more auditory perturbation when the after-effect is tested, and WP and PW were tested after WW and PP. This is a limitation of our experimental design, which will require further control by switching the order of same vs. other modalities in the protocol (that is WP then WW in group 1 and PW then PP in group 2). Interestingly, the after-effect seemed slightly lower in the WP condition than in the other conditions. Although this effect is not significant with the statistical power of the present study, it may pave the way to more extended work on this topic.

More generally, it is important to stress that even though the present results are not yet completely understood and in some cases inconclusive, they indicate that surface characteristics do matter in adaptation and transfer paradigms. Considering the increasing importance of this paradigm in recent research, the latter observation should be taken into account, especially in the design of experiments addressing adaptation or transfer in children or specific populations (such as illiterate adults).

5. Conclusions

Overall, the present observations display several effects that may be related to surface characteristics of stimuli. To the best of our knowledge, this study is the first to investigate modality effects in the transfer of auditory-motor adaptation. The results suggest that this topic could be insightful in the understanding of representations tied to speech articulation, and in particular to understand the level of abstraction of these representations.

6. Acknowledgements

The research leading to these results has received funding from the European Research Council under the European

Community's Seventh Framework Program (FP7/2007-2013 Grant Agreement no. 339152).

7. References

[1] L. Max and D. G. Maffett, "Feedback delays eliminate auditory-motor learning in speech production," *Neurosci. Lett.*, vol. 591, pp. 25–29, 2015.

[2] H. Terband, F. Van Brenk, and A. Van Doornik-Van Der Zee, "Auditory feedback perturbation in children with developmental speech sound disorders," *J. Commun. Disord.*, vol. 51, pp. 64–77, 2014.

[3] A. Daliri, E. A. Wieland, S. Cai, F. H. Guenther, and S. E. Chang, "Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter," *Dev. Sci.*, no. January 2016, pp. 1–11, 2017.

[4] J. F. Houde and M. I. Jordan, "Sensorimotor adaptation in speech production.," *Science*, vol. 279, no. 1998, pp. 1213–1216, 1998.

[5] E. N. MacDonald, E. K. Johnson, J. Forsythe, P. Plante, and K. G. Munhall, "Children's development of self-regulation in speech production," *Curr. Biol.*, vol. 22, pp. 113–117, 2012.

[6] S. D. Goldinger, "Echoes of echoes? An episodic theory of lexical access.," *Psychol. Rev.*, vol. 105, no. 2, pp. 251–279, 1998.

[7] H. L. Roediger and T. A. Blaxton, "Effects of varying modality, surface features, and retention interval on priming in word-fragment completion," *Mem. Cognit.*, vol. 15, no. 5, pp. 379–388, 1987.

[8] L. Ferrand, J. Grainger, and J. Segui, "A study of masked form priming in picture and word naming," *Mem. Cognit.*, vol. 22, no. 4, pp. 431–441, 1994.

[9] N. O. Schiller, "The effect of masked syllable primes on word and picture naming," *J. Acoust. Soc. Am.*, vol. 102, no. 5, p. 3136, 1997.

[10] D. W. Purcell and K. G. Munhall, "Compensation following real-time manipulation of formants in isolated vowels.," *J. Acoust. Soc. Am.*, vol. 119, no. 4, pp. 2288–2297, 2006.

[11] V. M. Villacorta, J. S. Perkell, and F. H. Guenther, "Sensorimotor adaptation to perturbations of vowel acoustics and its relation to perception," *J. Acoust. Soc. Am.*, vol. 122, no. 4, pp. 2306–2319, 2007.

[12] S. Cai, S. S. Ghosh, F. H. Guenther, and J. S. Perkell, "Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization.," *J. Acoust. Soc. Am.*, vol. 128, no. 4, pp. 2033–2048, 2010.

[13] A. Rochet-Capellan, L. Richer, and D. J. Ostry, "Nonhomogeneous transfer reveals specificity in speech motor learning," *J. Neurophysiol.*, vol. 107, no. 6, pp. 1711–1717, 2012.

[14] T. Caudrelier, P. Perrier, J. L. Schwartz, and A. Rochet-Capellan, "Does auditory-motor learning of speech transfer from the CV syllable to the CVCV word?," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2016, vol. 08–12–Sept, no. Cv, pp. 2095–2099.

[15] T. Caudrelier, J.-L. Schwartz, P. Perrier, S. Gerber, and A. Rochet-Capellan, "Transfer of learning: what does it tell us about speech production units?," *J. Speech, Lang. Hear. Res.*

[16] N. J. Bourguignon, S. R. Baum, and D. M. Shiller, "Lexical-perceptual integration influences sensorimotor adaptation in speech.," *Front. Hum. Neurosci.*, vol. 8, p. Article 208, 2014.

[17] A. F. Frank, "Integrating Linguistic, Motor, and Perceptual Information in Language Production," University of Rochester., 2011.

[18] A. Alsius, T. Mitsuya, N. Latif, and K. G. Munhall, "Linguistic initiation signals increase auditory feedback error correction," *J. Acoust. Soc. Am.*, vol. 142, no. 2, pp. 838–845, 2017.

[19] S. Cai, M. Boucek, S. S. Ghosh, F. H. Guenther, and J. S. Perkell, "A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/," in *Proceedings of the 8th ISSP*, 2008, pp. 65–68.

[20] R. Mundry and C. L. Nunn, "Stepwise model fitting and statistical inference: turning noise into signal pollution.," *Am. Nat.*, vol. 173, no. 1, pp. 119–123, 2009.

[21] T. Hothorn, F. Bretz, and P. Westfall, "Simultaneous inference in general parametric models," *Biometrical J.*, vol. 50, no. 3, pp. 346–363, 2008.

[22] J. W. Hawks, "Difference limens for formant patterns of vowel sounds.," *J. Acoust. Soc. Am.*, vol. 95, no. 2, pp. 1074–1084, 1994.