Second Language Speech Assimilation in an Optimal Transport Framework

Joselyn Rodriguez, Patrick Shafto, Naomi H. Feldman

1 Introduction

Some second language speech sounds are easier to learn than others. This variation is likely to be driven at least in part by the way second language speech sounds are related to native language categories (Strange, 2011; Best, 1995; Best and Tyler, 2007; Flege, 1995; Flege and Bohn, 2021; Chang, 2015; Escudero, 2005). Therefore, it is of great interest to understand how non-native listeners come to associate speech sounds in a second language to categories in the native language.

Although the mapping between non-native (L2) sounds and native language (L1) speech categories is often assumed to depend on acoustic or articulatory similarity (Best, 1995; Flege, 1995), second language learners do not always map second language sounds to L1 categories that are the most acoustically or articulatorily similar (Bohn, 2017; Strange, 2007, 1999). Alternative theories posit that learners know the phonological inventory of each language and match categories in the L1 to categories in the L2 based on phonological similarity (Chang, 2015; Best and Tyler, 2007; Strange, 2011). However, these theories do not typically make clear how this mapping takes into account the acoustic or articulatory properties of the different contrasts in each language, nor is there a mathematical account of how it could take place without reference to the these properties. Thus, although a range of theories tackle the problem of how learners map between L2 sounds to L1 categories, there is not yet a formal model that can predict how and why learners deviate from simply mapping each sound to its most acoustically or articulatorily similar category.

In this paper, we introduce a new framework for predicting the association between native language categories and second language speech sounds. Using the mathematics of optimal transport (Monge, 1781; Peyré and Cuturi, 2020), we derive a way of mapping between native speech categories and an acoustic

* Joselyn Rodriguez, Program for Neuroscience and Cognitive Science and Department Of Linguistics, University Of Maryland, jrodri20@umd.edu; Patrick Shafto, Department Of Mathematics and Computer Science, Rutgers University–Newark, patrick.shafto@gmail.com; Naomi H. Feldman, Department Of Linguistics and UMIACS, University Of Maryland, nhf@umd.edu. This research was supported by NSF grant BCS-2120834. We thank Bill Idsardi and the Acquisition Lab for helpful comments and discussion.

distribution[1] that formalizes the idea of an L1 filter. We find that when applied to second language speech sounds, this model correctly predicts mappings between L2 acoustics and native categories that go beyond acoustic similarity. Specifically, we apply the model to a case study of Spanish acquisition of English stops and show that the model's qualitative predictions are in line with what has been observed in second language learners. Our work provides insight into the optimization principles that could be guiding listeners as they transfer the native language phonological structure to novel L2 acoustic environments.

2 Spanish learners' acquisition of English stops

The case study we focus on in this paper is the acquisition of English stop contrasts by native speakers of Spanish. This scenario is illustrated in Figure 1. As can be seen from the figure, Spanish and English both maintain a two-way distinction along voice onset time (VOT) in word-initial positions. Inspecting an IPA chart would indicate that both languages both have a phonemic distinction between /d/ and /t/. However, the realization of these sounds differs phonetically across the two languages (Abramson and Lisker, 1973; Casillas et al., 2015; Williams, 1977b). English /d/ and /t/ are both realized as voiceless stops word-initially. The distinction between the two sounds in word initial position in English is actually aspiration: /d/ is realized as a voiceless stop [t] and /t/ is realized as [t^h]. In Spanish, /d/ is realized as pre-voiced whereas /t/ is voiceless, surfacing as [d] and [t] respectively (Figure 1; Abramson and Lisker 1973)[2].

Because both of the English categories are more acoustically similar to Spanish /t/ than to Spanish /d/, a second language learner that followed acoustic similarity alone would categorize both English sounds as Spanish /t/. However, this is suboptimal in second language learning because it collapses a contrast that is present phonologically in both the first and second language, and it is also not what is observed empirically. Rather, work on the acquisition of stops in Spanish-English bilinguals has found that Spanish-English bilinguals strive to maintain a two-way distinction (see Casillas 2021 for a review on production).

Beyond just retaining the distinction, previous work suggests that bilinguals are able to maintain representations that differentially process input from both English and Spanish. For example, studies examining the impact of language mode (Grosjean, 2020) for early and late bilinguals suggest language-specific patterns in discrimination (Lozano-Argüelles et al. 2021; Gonzales et al. 2019; Gonzales and Lotto 2013; García-Sierra et al. 2012; Elman et al. 1977; Casillas and Simonet

[1]We frame our simulations in terms of acoustic distributions, but the mathematical framework we develop could be applied to either acoustic or articulatory distributions. Our simulations in this paper are carried out in an idealized setting that is not derived directly from acoustic measurements of individual speech tokens, and thus do not provide new evidence in favor of acoustic representations.

[2]It has been noted in previous work that there is also a place distinction between the two languages: the Spanish sounds are produced as dental whereas the English sounds are produced as alveolar (Casillas et al., 2015). Given that most previous research has focused on VOT alone, we leave this for future work.

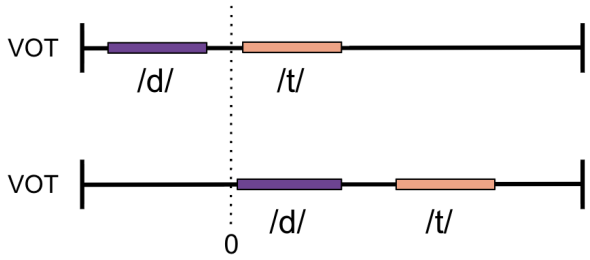

Figure 1: Categorization of /d/ and /t/ in English and Spanish. Spanish is "true voicing" language in which the voiced categories are pre-voiced and the voiceless category is short-lag. The English voiced category /d/, however, is realized as short-lag whereas the voiceless category /t/ is realized as long-lag.

2018; although see Bohn and Flege 1993; Caramazza et al. 1973; Williams 1977a for contradictory findings). Work focusing specifically on late L1 Spanish learners of English has found that these learners tend to perform similarly to the early bilinguals, displaying a shift in categorization boundaries towards the boundaries of monolingual English listeners along VOT (Gorba, 2019; Flege and Schmidt, 1995). Additional work utilizing a categorical oddity discrimination task has found native-like discrimination for the English /d/-/t/ ([t] - [t^h]) contrast for late L1 Spanish learners of English (Flege and Wayland, 2019)

These results suggest that learners are able to update category representations with exposure to a second language, and that this process takes place fairly quickly. More specifically, when learners acquire a second language with the same underlying contrast with different surface realizations, they update the representations to some extent in order to learn the L2-like categorization boundary. We hypothesize that this shift in the category boundary reflects the learners' mapping of their L1 categories onto new acoustic realizations in the L2. Here we formalize a computational model that optimizes a mapping between native categories and L2 sounds, and show that it can capture this type of shift in category boundary.

3 Model

We frame L2 language assimilation as a problem of optimal transport. The original formulation of optimal transport is as the solution to a straightforward problem: how to move piles of dirt to fill holes, minimizing the amount of effort (Monge, 1781). Since its initial formulation, optimal transport has found application across a variety of domains including machine learning (Arjovsky et al., 2017) and cognitive modeling (Wang et al., 2020) thanks to its general applicability to optimizing mappings between probability distributions.

To capture assimilation patterns in second language learners, we propose a model in which the speech sounds are analogous the dirt that is being moved and the categories are analogous to the holes. The effort needed to move a particular speech

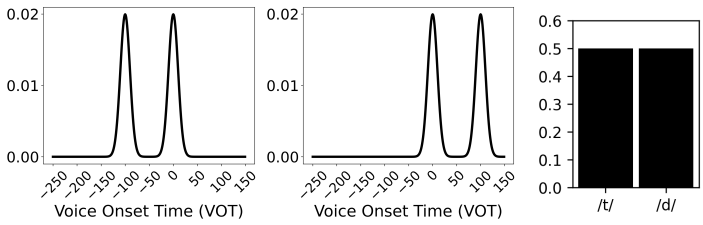

Figure 2: Left and middle figures: Distribution of acoustic information in two simulated languages. The Spanish-like native language simulates a true voicing language in which one category falls entirely in the negative value range while the English-like non-native language simulates an aspirating language with one category falling around 0 VOT and the other with a greater lag time. Right figure: Distribution over category labels. Each category is assigned .5 of the total probability mass.

sound into a particular category is defined in our model based on mappings between categories and acoustics in the native language. In acquiring a second language, a learner may encounter a new distribution of acoustics. The proposed model uncovers the optimal mapping between this encountered distribution of acoustics and the existing categories in the native language. Because optimal transport problems are constrained by both the dirt that is being moved and the holes that are being filled, our model's optimization is constrained by both the acoustics of the new language and the distribution over categories in the native language. The constraint that comes from the distribution over categories is important because it causes the categories from the native language to be redeployed in association with the distribution of acoustics in the second language. We show that the optimal solution to the resulting transport problem does not simply map sounds in the L2 to their most similar category in the L1. Instead, the model correctly predicts a boundary shift in the mapping between the acoustics and categories.

We compare our model to a Bayesian classifier that instantiates similarity-based assimilation (Best, 1995; Flege, 1995). Like the optimal transport model, the classifier operates over the acoustic distribution of sounds from the second language, but unlike the optimal transport model, it is not constrained to reproduce the distribution over categories in the first language. This allows us to isolate the effect of the category-based constraint in shaping assimilation patterns.

3.1 Background on Optimal Transport

Defining an optimal transport problem requires three parts: two distributions and a cost matrix. The solution to the problem is an *optimal transport plan*. This plan defines the most cost-efficient manner of coupling distributions. In this work, we use the discrete formulation of the problem.

Formally, a transport plan, P, is a joint probability matrix of size $m \times n$ that has the two probability distributions $\mu = (\mu_1 ..., \mu_m)$ and $\nu = (\nu_1 ..., \nu_n)$ as its marginals. In other words, μ is recoverable from P by summing over the separate values of ν and vice versa. There are many transport plans that have μ

and ν as their marginals, and the optimal transport plan minimizes the cost, C, of transporting between μ and ν (see Equation 1). In the current work, we make use of entropy regularized optimal transport which additionally includes a regularization parameter, $\epsilon > 0$, that enables efficient computation,

$$P^{\epsilon}(\mu, \nu) = \underset{P \in U(\mu,\nu)}{\operatorname{argmin}} \langle C, P \rangle - \epsilon H(P). \tag{1}$$

In Equation 1, μ and ν refer to two discrete finite probability distributions. $U(\mu, \nu)$ refers to the set of transport plans between μ and ν. H refers to the entropy of the distribution $H(P) = \sum_{i,j} -P_{i,j} \log P_{i,j}$ and $\epsilon \geq 0$. The cost matrix refers to a non-negative matrix, $C = (C_{ij})_{m \times n}$ that encodes the cost of transporting mass from a point i in one distribution to a point j in the other. The optimal transport plan, P^{ϵ}, is derived by minimization of Equation 1. The plan with the minimum entropy-regularized cost is the solution to the transport problem.

The solution to the problem can be computed using the Sinkhorn scaling algorithm, a method of deriving a balanced matrix that is able to return both the required marginal distributions (Knopp and Sinkhorn, 1967; Sinkhorn, 1964). Importantly, this algorithm is able to solve the optimal transport problem in Equation 1. Sinkhorn scaling is completed by first initializing a candidate plan using a Gibbs kernel: $K_{i,j} = e^{-C_{ij}/\epsilon}$. The algorithm then iteratively refines the initial guess by alternating projection onto each marginal distribution. When changes are small enough, we terminate computation. We implement this algorithm using the POT: Python Optimal Transport package (Flamary et al., 2021) using a value of $\epsilon = 1$.

3.2 Assimilation as Optimal Transport

We now define the our optimal transport model of L2 assimilation. The two probability distributions that we utilize are a distribution over acoustics, $P(X)$ (Figure 2 left) and a distribution over category labels, $P(Y)$ (Figure 2 right). The discretized distributions are the marginals distributions, μ and ν, utilized in the optimal transport problem (Equation 1). Thus, the output of this model is the optimal map between acoustic values (here, an idealized version of VOT) and category labels (/d/ or /t/).

The distribution over acoustics is constructed using two normalized Gaussian distributions. We simulate the native language by creating two Gaussian distributions centered around -100ms and 0ms with standard deviations of 10ms, normalized to sum to 1 (Figure 2 left). This language serves as the simulated Spanish-like native language. The second language (long-lag or English-like) is similarly constructed with two Gaussians with means centered at 0ms and 100ms each with standard deviations of 10ms (Figure 2 middle). This serves as the English-like L2 language. Each distribution is discretized by binning over the VOT values, so that the distribution over acoustics is a discrete distribution defined over 400 VOT bins ($m = 400$). The distribution over category labels consists of a uniform discrete distribution with two values /t/ and /d/ ($n = 2$; Figure 2 right).

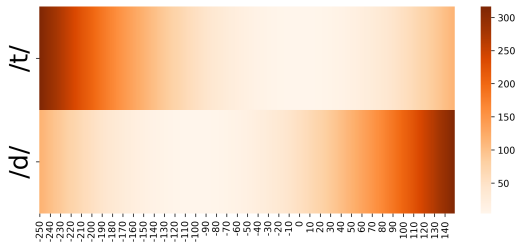

Figure 3: Cost matrix derived as the negative log probability of the joint distribution between L1 acoustic distribution and category distribution. Lighter colors correspond to lower cost values and thus higher probability.

The last piece is the cost matrix, which we define in terms of the negative log of the joint probability distribution between acoustics, X, and category labels, Y, derived from the native language,

$$C = -\log P(X, Y). \tag{2}$$

For each native language category (/t/ or /d/), the cost matrix assigns low cost to the acoustics for the category dependent on the acoustic realization of that category in the native language. These values are computed from the simulated acoustic distributions for each category in the native language, resulting in a $m \times n$ matrix (Figure 3)[3].

With all the pieces in place, we now compute the optimal transport plan via Equation 1 implemented using the Sinkhorn scaling algorithm (Sinkhorn, 1964). Recall this takes place through iterative refinement of an initial guess K. In this calculation, because the cost matrix is defined in terms of the log probability function, the matrix K that initializes the Sinkhorn scaling algorithm reduces to $P(X, Y)$, the joint probability distribution over acoustics and category labels in the native language. In each step, the algorithm iteratively derives the optimal plan by projecting onto one marginal distribution and normalizing over the other to satisfy the constraint over marginals for transport plans.

3.3 Assimilation as Bayesian classification

To illustrate the utility of the proposed model, we additionally implement a baseline model: a Bayesian classifier, which has been used as a baseline model for computational modeling of L1-L2 perception in previous work (Escudero et al., 2007). A Bayesian classifier is a statistical model that makes use of Bayes' rule to compute posterior probabilities $P(Y|X)$ of category membership,

$$P(Y|X) = \frac{P(X|Y)P(Y)}{\sum_{Y'} P(X|Y')P(Y')} \tag{3}$$

[3]The cost matrix, as well as the transport plans in Figure 4, are transposed for easier visualization, with μ on the columns and ν on the rows.

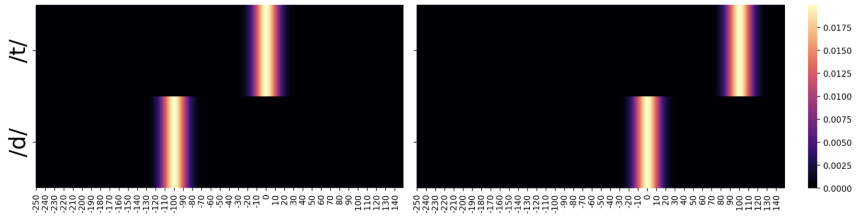

Figure 4: Optimal transport plan for acoustic values in L1 distribution to L1 categories (left) and for L2 acoustics and L2 categories (right). The y-axis are the category labels, /d/ and /t/, the lighter shades show the values from the acoustic distribution that the model associates strongly with a category label.

where the likelihood, $P(X|Y)$, is a Gaussian distribution corresponding to a particular category label in the native language, i.e. one of the Gaussians illustrated in Figure 2 (left), and the prior distribution, $P(Y)$, is the distribution over category labels illustrated in Figure 2 (right).

This has connection to the proposed model as it can be understood as one part of the transport problem, namely marginalization along the acoustics. This is because the cost matrix C and initialization matrix K correspond to the joint distribution over acoustics and category labels—the numerator in Equation 3—and because the normalizing constant in Equation 3 is equal to $P(X)$. However, the Bayesian model differs from the optimal transport model in that it doesn't require the marginalization over the category labels. Intuitively, this can be understood as maintaining a native-language filter through the categorization model, but importantly, not constraining the filter to maintain any particular distribution over category labels.

4 Simulations

We present two simulations to illustrate our model's behavior. Simulation 1 shows that both our optimal transport model and the baseline model accurately recover the association between categories and labels in the native language when presented with samples from the same acoustic distribution that they were trained on. Simulation 2 shows that when the distribution of acoustics is changed to that of a second language, only the optimal transport model is a able to predict a shift in the mappings between acoustics and category labels.

4.1 Simulation 1: Native Language Assimilation

The goal of Simulation 1 is to show that when the model encounters a native-like distribution over acoustics, it recovers the same mappings between categories and acoustics that it was trained on.

We first discuss the optimal transport model. We set up the model in terms of the native speaker (Spanish-like). We use the cost matrix derived via the native-like

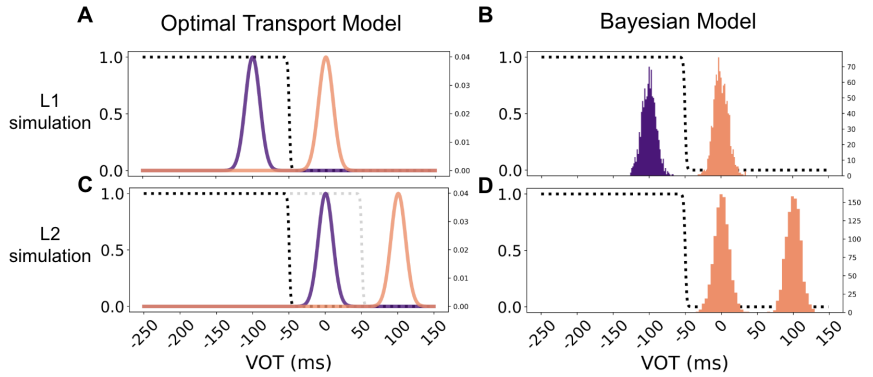

Figure 5: Categorization of the optimal transport and Bayesian model when trained on the L1 and tested on the L1 (A,B) and L2 (C,D). Purple corresponds to the /d/ label, and orange corresponds to the /t/ label. The logistic function indicates the location of the category boundary in the native language.

patterns as discussed in Section 3.1. In this simulation, the marginal distribution of acoustic values (VOT) and category labels are exactly the same distributions that this cost matrix was derived from; they are taken from previous work on the distribution of VOT and perceptual behavior by native speakers (Abramson and Lisker, 1973). Thus, we solve the optimal transport problem between the unlabeled distribution of acoustics (Figure 2 left) and category labels (Figure 2 right).

Because the cost matrix is derived from the native language, the optimal transport plan simply recovers the L1 mapping between acoustic values and categories. This can be seen in the left of Figure 4, in which the lighter shaded colors are at the same locations along the x-axis (VOT) as in the original distribution (Figure 2 left), and in the correct boundary in Figure 5a. This result is expected given the formulation of the cost matrix directly from the native language data. Nevertheless, it is worthwhile to note that this model is able to use the cost matrix that was derived from the native language to correctly map a distribution of *unlabeled* VOT values to their corresponding labels.

The Bayesian model is tested on native language (Spanish-like) acoustic values drawn from the distribution in Figure 2 (left). We find that the Bayesian model also derives the correct pattern (Figure 5b) for the native language speakers. This is also an expected result as the model's prior and likelihood encode the correct joint distribution over acoustics and categories in the native language, and thus yield an accurate posterior for categorizing each sound. Therefore both the optimal transport model and Bayesian classifier correctly predict patterns of categorization for the native language listeners.

4.1.1 Simulation 2: L2 Assimilation

The goal of Simulation 2 is to extend the models to the L2 learning scenario. To behave like human second language learners, the model should be able to shift the category boundary to account for a novel distribution of acoustics.

Again, we first discuss the optimal transport model. The L2 simulation differs from the native simulation *only* in the input acoustic distribution, μ. Rather than deriving a transport plan for the acoustic distribution of the native language (left of Figure 2), it utilizes the acoustic distribution of the L2 (middle). The distribution over category labels and the cost matrix remain the same. Intuitively, this can be understood as maintaining the same phonemic categories and native-language filter, but receiving an unfamiliar acoustic distribution of VOT values. Of main interest to the current work is the application of this L1-derived cost matrix to the *novel* distribution of VOT from the L2 (Figure 2 middle).

We find that when the optimal plan is derived for the L2 acoustic distribution (English-like) via the L1-derived cost matrix, the result is a transport plan that shifts category labels to accommodate the novel distribution, such that the Gaussian distribution around 0ms is now mapped to /d/ whereas the Gaussian distribution around 100ms is mapped to /t/ (Figure 4 right and Figure 5c). As can be seen from Figure 5 which shows both the native and non-native transport plans, the distribution of acoustics is mapped not according to the phonetically most similar category, but rather the overall distribution over category labels. Thus, VOT values are mapped to both categories.

We now discuss the results for the Bayesian model. We find that the model overgeneralizes the voiceless category, /t/ label (Figure 5d). The model overgeneralizes because its likelihood is based on the native language. The L2 boundary shift cannot be achieved from the Bayesian classification model since there is no constraint on the marginal distribution over categories. In effect, all the VOT values are moved to only one category, the category that is phonetically most similar.

The difference in the models' categorization behavior arises despite the fact that both the Bayesian and the optimal transport model are trained on distributions derived from the native language. The optimal transport model arrives at the human-like solution because it constrains the transport plan to maintain the marginal distributions over both the acoustics and category labels. The result of this constraint is that acoustic values must be assigned to both of the labels. The optimal plan results in a shift in categorization boundary along VOT rather than assigning all of the VOT values to a single category. The Bayesian model has no such constraint that requires the data to be distributed over each label, resulting in the collapse to the closest L1 category.

5 Discussion

We introduced a model that formalizes second language speech assimilation using optimal transport. Through two simulations, we show that the optimal transport

model is able to account for initial categorization patterns in L1 *and* subsequent shifts in L2 utilizing a single filter (i.e., cost matrix) derived through the L1. The result of L2 mapping can thus be explained as the result of an optimization problem that aims to find the best coupling between categories in the native language and the acoustics of the second language.

Our model differs from previous accounts in its focus on deriving the optimal mapping between acoustics and categories (but see Escudero, 2005). It is able to predict L2 assimilation patterns that previous accounts have assumed a priori. The Perceptual Assimilation Model (PAM) (Best, 1995) and its L2 extension PAM-L2 (Best and Tyler, 2007) aims to explain discrimination patterns in perception through assimilation patterns. L2 contrasts are assimilated to native language categories dependent on articulatory similarity between sounds in the contrasts. The theory does not account for how this metric of similarity is determined, but rather relies on the outcome of assimilation studies with humans as the basis for predicting discrimination. The Speech Learning Model (SLM) (Flege, 1995; Flege and Bohn, 2021) also claims that the difficulty of perceiving L2 speech sounds is dependent on phonetic similarity to L1 categories, but does not provide a method through which to determine how these equivalences may take place. The current model addresses this gap in previous theories by modeling assimilation as the optimal result of mapping L2 acoustics to native categories.

The model also differs from previous theories that appeal to phonological similarity (Best and Tyler, 2007; Chang, 2015) in that it bypasses the phonological structure of the L2 entirely, and imposes the categories of the native language on the acoustics of the L2. This aspect of our model may need to be revised in future work. For example, it may be that learners derive the category-based optimal transport constraint from their L2 rather than from their native language, or that they only force the native category structure onto their L2 once they know that there is a relevant contrast in their L2 that they should map their L1 categories onto. Importantly, our model also explicitly requires the maintenance of two categories in perception in both the L1 and L2 scenarios. Depending on how these categories are defined, this could suggest that phonological information in the native language, such as the structure of categories along a dimension (VOT), plays a role in perception of non-native languages.

Finally, the proposed model differs from previous accounts of perceptual adaptation in which supervised training led to shifts in category boundaries (Kleinschmidt and Jaeger, 2015). This difference is crucial because supervised training requires a learner to already know the correct mapping between categories and acoustics. Our model derives this mapping from first principles, which allows it to capture a shift in the category boundary without requiring labeled training data in the L2.

There are several questions that remain. As we discussed before, in the current work and the broader literature, it's unclear what the correct level of abstraction is when considering the influence of the native language on L2 perception (Chang, 2015; Best and Tyler, 2007; Broselow and Kang, in press; Bohn, 2017) or whether this should take place along an acoustic, articulatory, or featural basis (Yazawa et al.,

2023; Brown, 1998; Archibald, 2023). In the current model, we do not specify whether the L1 categories under discussion are phonetic or phonological in nature. Given that the acoustic distributions used in our simulations are intended to mimic those that occur in word-initial position, either interpretation is possible. However, given the flexibility of the optimal transport framework, future simulations need not be restricted to considering only the sounds that occur in a single context. For example, one could derive a cost matrix based on more complex distributions in which allophonic variation is mapped to a single phonological category in the native language (e.g., Rohena-Madrazo, 2013). More generally, to tackle these questions, it is desirable to have a way to compare second language learners' predicted assimilation patterns under different modeling assumptions. The optimal transport framework we have introduced can facilitate this type of comparison because it provides the precision of mathematical optimization while allowing considerable flexibility in specifying the relevant probability distributions.

The proposed model provides an initial formalization of speech assimilation in a second language. While the model captures the expected pattern, several simplifications were made that should be addressed in future work. For example, the Gaussian distributions for each category were assumed to have the same frequency and variance across languages; however, generally, this doesn't prove to be true (Lisker and Abramson, 1964). Additionally, we only focus on the situation in which the number of categories across languages is the same. How listeners may deal with differing numbers of categories cross-linguistically, and the situations in which categories must be created is an interesting future direction. Optimal transport provides a promising tool for exploring cases in which learners need to create new categories because it explicitly assigns a cost for transport between acoustics and categories. This could provide a concrete method of quantifying predicted cost of assimilation to native language categories relative to creation of new categories altogether. More broadly, the cost assigned for mapping a particular set of native language categories to an acoustic distribution in a second language can be explored in future work as a way of predicting the difficulty of a particular second language learning scenario.

References

Abramson, Arthur S., and Leigh Lisker. 1973. Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics* 1:1–8. Publisher: Elsevier BV.

Archibald, John. 2023. Phonological redeployment and the mapping problem: Cross-linguistic E-similarity is the beginning of the story, not the end. *Second Language Research* 39:287–297. Publisher: SAGE Publications Ltd.

Arjovsky, Martin, Soumith Chintala, and Léon Bottou. 2017. Wasserstein GAN. URL `http://arxiv.org/abs/1701.07875`, arXiv:1701.07875 [cs, stat].

Best, Catherine. 1995. A direct realist view of crosslanugage speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 171–204. Journal Abbreviation: Speech Perception and Linguistic Experience: Issues in Cross-Language Research.

Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Language Learning & Language Teaching*, ed. Ocke-Schwen Bohn and Murray J. Munro, volume 17, 13–34. Amsterdam: John Benjamins Publishing Company.

Bohn, Ocke-Schwen. 2017. *Cross-language and second language speech perception*, chapter 10, 213–239. John Wiley Sons, Ltd.

Bohn, Ocke-Schwen, and James Emil Flege. 1993. Perceptual switching in Spanish/English bilinguals. *Journal of Phonetics* 21:267–290.

Broselow, Ellen, and Yoonjung Kang. in press. Phonology and phonetics. In *The cambridge handbook of second language acquisition*. Cambridge University Press, 2 edition.

Brown, C.a. 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14:136–193.

Caramazza, A., G. H. Yeni-Komshian, E. B. Zurif, and E. Carbone. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America* 54:421–428.

Casillas, Joseph V. 2021. Interlingual Interactions Elicit Performance Mismatches Not "Compromise" Categories in Early Bilinguals: Evidence from Meta-Analysis and Coronal Stops. *Languages* 6:9.

Casillas, Joseph V, Yamile Díaz, and Miquel Simonet. 2015. Acoustics of spanish and english coronal stops. In *ICPhS*.

Casillas, Joseph V., and Miquel Simonet. 2018. Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics* 71:51–64.

Chang, Charles. 2015. Determining cross-linguistic phonological similarity between segments: The primacy of abstract aspects of similarity. In *The Segment in Phonetics and Phonology*, 199–217. Journal Abbreviation: The Segment in Phonetics and Phonology.

Elman, Jeffrey L., Randy L. Diehl, and Susan E. Buchwald. 1977. Perceptual switching in bilinguals. *The Journal of the Acoustical Society of America* 62:971–974.

Escudero, Paola, Jelle Kastelein, Klara Weiand, and R. J. J. H. Van Son. 2007. Formal modelling of L1 and L2 perceptual learning: computational linguistics versus machine learning. In *Interspeech 2007*, 1889–1892. ISCA.

Escudero, Paola Rocío. 2005. *Linguistic perception and second language acquisition: explaining the attainment of optimal phonological categorization*. Number 113 in LOT. Utrecht: LOT.

Flamary, Rémi, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. 2021. Pot: Python optimal transport. *Journal of Machine Learning Research* 22:1–8.

Flege, James, and Anna Schmidt. 1995. Native Speakers of Spanish Show Rate-Dependent Processing of English Stop Consonants. *Phonetica* 52:90–111.

Flege, James E. 1995. Second language speech learning: Theory, findings and problems. In *Speech perception and linguistic experience: Issues in cross-language research*, 233–277. York Press.

Flege, James E., and Ratree Wayland. 2019. The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies* 2:1–44.

Flege, James Emil, and Ocke-Schwen Bohn. 2021. The Revised Speech Learning Model (SLM-r). In *Second Language Speech Learning*, ed. Ratree Wayland, 3–83. Cambridge University Press, 1 edition.

García-Sierra, Adrián, Nairán Ramírez-Esparza, Juan Silva-Pereyra, Jennifer Siard, and Craig A. Champlin. 2012. Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study. *Brain and Language* 121:194–205.

Gonzales, Kalim, Krista Byers-Heinlein, and Andrew J. Lotto. 2019. How bilinguals perceive speech depends on which language they think they're hearing. *Cognition* 182:318–330.

Gonzales, Kalim, and Andrew J. Lotto. 2013. A bafri, un pafri: Bilinguals' pseudoword identifications support language-specific phonetic systems. *Psychological Science* 24:2135–2142.

Gorba, Celia. 2019. Bidirectional influence on L1 Spanish and L2 English stop perception: The role of L2 experience. *The Journal of the Acoustical Society of America* 145:EL587–EL592.

Grosjean, François. 2020. The bilingual's language modes 1. In *The bilingualism reader*, 428–449. Routledge.

Kleinschmidt, Dave F., and T. Florian Jaeger. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review* 122:148–203.

Knopp, Paul, and Richard Sinkhorn. 1967. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics* 21:343–348. Publisher: Pacific Journal of Mathematics, A Non-profit Corporation.

Lisker, Leigh, and Arthur S. Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD* 20:384–422. Publisher: Routledge _eprint: https://doi.org/10.1080/00437956.1964.11659830.

Lozano-Argüelles, Cristina, Laura Fernández Arroyo, Nicole Rodríguez, Ezequiel M. Durand López, Juan J. Garrido Pozú, Jennifer Markovits, Jessica P. Varela, Núria De Rocafiguera, and Joseph V. Casillas. 2021. CONCEPTUALLY CUED PERCEPTUAL CATEGORIZATION IN ADULT L2 LEARNERS. *Studies in Second Language Acquisition* 43:204–219.

Monge, Gaspard. 1781. Memory on the theory of excavations and embankments. *History of the Royal Academy of Sciences of Paris* .

Peyré, Gabriel, and Marco Cuturi. 2020. Computational Optimal Transport. ArXiv:1803.00567.

Rohena-Madrazo, Marcos. 2013. Perceptual assimilation of occluded voiced stops by spanish listeners. In *Selected Proceedings of the 15th Hispanic Linguistics Symposium*, 140–156.

Sinkhorn, Richard. 1964. A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices. *The Annals of Mathematical Statistics* 35:876–879. Publisher: Institute of Mathematical Statistics.

Strange, Winifred. 1999. Levels of abstraction in characterizing cross-language phonetic similarity. In *Proceedings of the 14th international congress of phonetic sciences*, 2513–2519.

Strange, Winifred. 2007. Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In *Language Learning & Language Teaching*, ed. Ocke-Schwen Bohn and Murray J. Munro, volume 17, 35–55. Amsterdam: John Benjamins Publishing

Company.
Strange, Winifred. 2011. Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics* 39:456–466.
Wang, Pei, Junqi Wang, Pushpi Paranamana, and Patrick Shafto. 2020. A mathematical theory of cooperative communication. In *Advances in Neural Information Processing Systems*, volume 33, 17582–17593. Curran Associates, Inc.
Williams, Lee. 1977a. The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics* 21:289–297.
Williams, Lee. 1977b. The voicing contrast in Spanish. *Journal of Phonetics* 5:169–184.
Yazawa, Kakeru, James Whang, Mariko Kondo, and Paola Escudero. 2023. Feature-driven new sound category formation: computational implementation with the L2LP model and beyond. *Frontiers in Language Sciences* 2. Publisher: Frontiers.