Applied Psycholinguistics (2026), **47**, e58, 1–26
doi:10.1017/S0142716426101076

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Modeling cross-language vowel similarity: integrating phonetic analysis and explainable classification

Georgios P. Georgiou and Aretousa Giannakou

Department of Languages and Literature, University of Nicosia, Cyprus
Corresponding author: Georgios P. Georgiou; Email: georgiou.georg@unic.ac.cy

(Received 25 January 2026 UTC; revised 24 August 2026 UTC; accepted 15 September 2026 UTC)

Abstract

This study provides a multidimensional comparison of Standard Modern Greek and Standard Albanian, examines bidirectional cross-language vowel categorization, and identifies the acoustic cues that drive discriminability. Forty-two adults produced nonce words in carrier phrases; we extracted midpoint frequencies for the first three formants, formant dynamics, and vowel duration, and analyzed acoustic differences with mixed-effects models. We also simulated bidirectional category mapping using Random Forest classifiers trained on one language and applied to the other, and used explainable artificial intelligence (XAI) to identify the features driving classification. The results showed that the shared five-vowel inventory was comparable across languages in overall spatial overlap, though with selective spectral/dynamic differences and pervasive timing divergence. The mapping simulations yielded largely corresponding mappings for shared vowels in both directions, although the reciprocal mappings were not fully symmetrical. Albanian /ə/ was classified mainly as Greek /o/, and Albanian /y/ primarily as Greek /i/. The machine learning classifier performed well, with XAI indicating duration as the dominant cue. The study demonstrates that cross-language vowel similarity is inherently multi-dimensional, since apparent equivalence in canonical vowel space can coexist with systematic temporal and spectral divergence that supports robust discriminability and predictable category mapping.

Keywords: Albanian; cues; explainability; Greek; machine learning; vowels

Introduction

Vowels constitute a core component of phonological contrast and lexical identity across the world's languages, and they are among the most intensively studied segments in acoustic phonetics (e.g., see Adem, 2026; Georgiou & Giannakou, 2024; Munro, 2022; Song & Eckman, 2025). A long tradition of work has shown that vowel quality is reliably captured by formant structure, especially F1 (height) and F2

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

(frontness), while additional spectral information (e.g., F3) and temporal properties (e.g., duration) often provide crucial secondary cues for category identification and contrast maintenance (Hillenbrand et al., 1995; Peterson & Barney, 1952). Importantly, vowels are not purely static targets: formant trajectories can carry language-specific structure (vowel-inherent spectral change), and listeners can exploit dynamic information when vowel categories overlap in midpoint space or when coarticulation is substantial. Consequently, a complete account of vowel system similarity and difference must consider multiple dimensions of the signal—including static formant measures, dynamic measures derived from formant movement, and timing-related cues (Elvin et al., 2021; Williams et al., 2018). These questions can be addressed not only through direct acoustic comparison of native productions, but also through computational simulations of category mapping, in which vowels from one language are classified relative to the category structure of another.

From a typological standpoint, vowel inventories display both common structural tendencies and language-particular implementations. Five-vowel systems of the type /i, e, a, o, u/ are widespread (Maddieson, 1984) and often described as maximally dispersed solutions under articulatory–acoustic constraints (Liljencrants & Lindblom, 1972). However, languages that share the same inventory can still differ systematically in phonetic realization, e.g., shifts in where /a/ or /o/ sit in F1–F2 space and in how tightly categories are spaced, reflecting language-specific patterns of implementation and dispersion (Recasens & Espinosa, 2009). Moreover, vowel contrasts are frequently supported by temporal and dynamic cues in addition to midpoint formant values, with formant movement and other time-varying properties contributing to robust identification and contrast maintenance (Nittrouer, 2007). These phonetic differences shape cross-language perceptual similarity and categorization: even "shared" phonological categories may not be straightforwardly equivalent in production and perception, as shown in assimilation-based accounts and in work linking acoustic similarity to cross-language vowel perception (Best & Tyler, 2007; Georgiou & Dimitriou, 2023; Williams & Escudero, 2014). For instance, Georgiou and Dimitriou (2023) showed that phonological "overlap" does not guarantee categorical equivalence: although Dutch and Cypriot Greek both include /e/ and /o/, Dutch productions were mapped onto the corresponding Cypriot Greek categories only 67% (/e/) and 68% (/o/) of the time, indicating substantial cross-language ambiguity in category alignment.

Greek and Albanian offer a particularly informative comparison. Standard Modern Greek is commonly described as having the prototypical five-vowel inventory /i, e, a, o, u/ (Arvaniti, 1999; Topintzi, 2023). Standard Albanian includes the same set of five vowels but also contains additional categories; most notably /ə/ and /y/ (Coretta et al., 2023), which expand the system and may influence the realization of the shared vowels through inventory-wide dispersion and contrast maintenance. Comparing a five-vowel system (Greek) to a seven-vowel system (Albanian) is typologically consequential because inventory size can affect how vowels are distributed in acoustic space and, in some accounts, how contrasts are structured at the system level (Al-Tamimi & Ferragne, 2005; Schwartz et al., 1997). If Albanian's additional categories compress the vowel space or carve out intermediate regions, such as a central vowel (/ə/) and the high front rounded vowel (/y/), then

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

shared vowels may show subtle shifts in F1–F2 location and potentially in other dimensions implicated in contrast maintenance. Moreover, because both languages are spoken in the Balkan region, a well-known linguistic area characterized by long-term multilingual contact and convergence, Greek–Albanian comparisons can also inform broader questions about how vowel systems remain stable or diverge under sustained geographic proximity and contact (Friedman & Joseph, 2017).

Empirical work on Greek–Albanian bilingual and L2 speech provides evidence that the phonological similarity of the two vowel systems does not necessarily result in fully native-like phonetic realization. Kaskampa and Georgiou (2025), examining Greek vowels produced by adult Albanian speakers of Greek as an L2, found differences from L1 Greek productions for all five vowels on at least one acoustic dimension, although several individual formant and durational properties were produced in a native-like manner. The authors attributed this combination of similarity and divergence partly to the phonological correspondence between the two vowel inventories and partly to L1–L2 interference arising from phonetic differences between corresponding categories. Related evidence comes from Albanian heritage speakers dominant in Greek, whose Greek vowels differed from those of monolingual Greek speakers in selected formant measures and, notably, in the duration of all five vowels (Georgiou & Giannakou, 2024). These findings indicate that extensive exposure to Greek and the existence of corresponding phonological categories do not necessarily eliminate fine-grained cross-language phonetic differences. Direct research on the perceptual assimilation of Greek and Albanian vowels remains comparatively limited, however, making an acoustically grounded examination of the relationship between the two systems particularly useful for generating predictions about cross-language categorization.

Although the present study focuses on vowel production in monolingual speakers rather than on perception or L2 learning, cross-language speech-perception models provide a useful theoretical basis for examining similarity, category mapping, and cue weighting in vowel systems. The Speech Learning Model and its revised form (SLM-r) argue that L2 sound learning is shaped by the relationship between new sounds and established L1 categories, with category formation being less likely when nonnative sounds are perceived as highly similar to existing L1 categories (Flege & Bohn, 2021). Similarly, PAM and PAM-L2 propose that nonnative phones are perceived relative to native phonological categories, and that discrimination depends on assimilation patterns such as two-category or single-category assimilation (Best & Tyler, 2007). The L2LP model adds a computationally explicit account by linking perception and learning to acoustic distributions and perceptual boundary formation (Van Leussen & Escudero, 2015). More recently, the Universal Perceptual Model (UPM) has emphasized that nonnative perception is shaped by native-language filtering and by cross-linguistic overlap in acoustic–perceptual space, while also treating acoustic and perceptual similarity as distinct but interacting predictors of categorization and discrimination (Georgiou, 2021; Georgiou & Dimitriou, 2023). UPM further predicts that linguistic experience, including bidialectalism, can modulate sensitivity to nonnative contrasts by reshaping category structure and cue weighting (Georgiou et al., 2025). Related production-oriented work has shown that overlap metrics based on formants and

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

duration can illuminate category proximity and potential confusability in L2 vowel productions (Georgiou & Savva, 2026).

These constructs are also relevant to typological phonetic comparison. If two languages share vowel categories, it is possible to ask how tokens from one language would be categorized relative to the category structure of the other. In perception research, this question is usually addressed through listener classification tasks; however, computational simulation offers a complementary approach when perceptual data are unavailable or when the aim is to model acoustic category mapping directly. Supervised classifiers can be trained on one language and then used to predict how tokens from another language map onto the trained categories, thereby providing quantitative evidence about cross-language similarity and discriminability (Alispahic et al., 2017; Georgiou, 2023; Georgiou, 2024). Machine learning methods are useful in this context because they can integrate multiple acoustic predictors, including F1–F3, trajectory-derived measures, and duration, while allowing nonlinear decision boundaries and out-of-sample evaluation (Themistocleous, 2019). For example, Georgiou (2025a) showed that F1, F2, F3, and duration could reliably distinguish L1 English speakers from Cypriot Greek speakers of English as an L2, demonstrating that these measures capture systematic cross-language variation in vowel realization. Random Forests are particularly suitable because they are robust to correlated predictors, capture complex interactions, and perform well without strong distributional assumptions; their use of bootstrap aggregation and random predictor selection helps reduce variance and prevents correlated variables from consistently dominating model decisions (Breiman, 2001).

Finally, because prediction alone does not explain which phonetic dimensions drive classification, the present study incorporates explainable artificial intelligence (XAI) (Dwivedi et al., 2023; Georgiou, 2025b). XAI methods identify which input features contribute most strongly to model predictions, allowing classification results to be interpreted in terms of phonetic cue use rather than accuracy alone. Shapley additive explanations (SHAP) provides a principled feature-attribution framework for quantifying predictor contributions at both token-specific and aggregate levels (Lundberg & Lee, 2017). Recent speech research has used SHAP to identify influential acoustic cues in classification models; for instance, Benway et al. (2024) found that F3 was the strongest predictor of fully rhotic productions in children with speech sound disorders, while Georgiou and Paphiti (2026) reported that F0, intensity, F3, and F1 were the most important predictors in classifying autistic and neurotypical adults. Thus, combining supervised classification with SHAP makes it possible to examine not only whether Greek and Albanian vowels differ, but also which acoustic dimensions—static formants, dynamic measures, or duration—most strongly structure model-based cross-language discriminability.

This study

This study has three main aims. First, it seeks to investigate typologically informed phonetic differences between Greek and Albanian vowels, with a particular focus on vowels that are shared by the two languages. Second, it employs computational machine learning methods to simulate the classification of Albanian vowels in terms of Greek vowel categories, with an additional reverse classification of Greek vowels

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

in terms of Albanian vowel categories, and to quantify the discriminability of the shared vowel inventory on the basis of detailed acoustic features. Third, it aims to identify which acoustic features contribute most strongly to vowel classification by examining feature importance within the machine learning classification model.

The originality and contribution of this study are threefold. Empirically, it offers a systematic, acoustically grounded comparison of Greek and Albanian vowels that jointly examines static, dynamic, and temporal measures, rather than restricting analysis to midpoint F1–F2 alone. This matters because cross-language differences may be subtle in vowel-space location yet substantial in duration or in trajectory structure, properties that are increasingly recognized as integral to vowel contrasts and cross-language similarity. Methodologically, the study extends computational category-mapping approaches, which were previously employed primarily in L2 classification contexts, to a typological, cross-language setting, providing a quantitative simulation of how one language's vowel tokens align with another language's vowel categories. In doing so, it complements traditional mixed-effects inference with a discriminability-based perspective grounded in supervised learning. Finally, analytically, the study uses SHAP-based interpretability to translate classifier behavior into phonetic insight: it identifies which features the model treats as most diagnostic and, therefore, which dimensions of the acoustic signal most strongly differentiate Greek and Albanian vowel realizations. This interpretability step directly informs theoretical discussion about cue weighting and category similarity as framed by speech models, and it provides a replicable template for future cross-language and contact-phonetics work.

Methodology

Participants

The study included 42 adult participants. Twenty-one were native speakers of Standard Modern Greek (11 females), aged 21–53 years ($M = 35.5$, $SD = 10.2$), born and raised in Greece and residing permanently in Athens. An additional 21 participants were native speakers of Standard Albanian (12 females), aged 18–54 years ($M = 34.24$, $SD = 11.47$). Because the study focused on adult L1 speakers, the groups had comparable age distributions, and no children or elderly speakers were included; age was not included as a covariate in the analyses. All participants were born and raised in Albania and resided permanently in central and central-southern Albania. Across both groups, participants reported a moderate socioeconomic background, no extended residence abroad, no use of additional languages or dialects, and no history of language, cognitive, or hearing impairment.

Materials

The stimulus set comprised five monosyllabic nonsense words in an /fVf/ frame, designed to elicit each of the five Greek (/i, e, a, o, u/) vowels and each of the seven Albanian vowels (/i, e, a, o, u, ə, y/), respectively. The target items were embedded in carrier phrases: Greek "Ípes /fVf/ páli" and Albanian "E the /fVf/ prap" ("You said /fVf/ again"). The choice of the /fVf/ context was motivated by experimental control rather than lexical availability. We selected a symmetrical voiceless fricative context to

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

provide a consistent phonetic environment across all vowels and both languages while minimizing contextual variability. Using the same voiceless fricative before and after the vowel allowed us to maintain a uniform consonantal context across tokens and languages and to facilitate direct cross-language comparisons of vowel acoustics.

Procedure

Each participant completed the task individually in quiet rooms. The sentences were presented on paper, and the researcher instructed participants to read them as they would when speaking to a friend. The materials were written in Standard Modern Greek and Standard Albanian orthographies. Speech was recorded with a professional recorder at a 44.1 kHz sampling rate and saved as 24-bit Waveform files. The Greek participants produced 10 utterances (5 vowels × 2 repetitions), and the Albanian participants produced 14 utterances (7 vowels × 2 repetitions). Before recording, participants completed a brief practice run in which they read all sentences aloud; the researcher monitored performance and clarified any misread words or pronunciations before the task began. The target sentences were presented in a fixed order for all participants rather than being randomized. This was done to maintain procedural consistency across speakers and languages. Given the small number of stimuli and the use of a brief practice run before recording, potential order effects were expected to be limited.

The target words were segmented and analyzed in Praat (Boersma & Weenink, 2025). Spectrograms and waveforms were inspected visually to identify relevant acoustic landmarks and to establish vowel boundaries for measuring formant frequencies and vowel duration. Analyses were conducted using a 0.025 s analysis window, 50 Hz pre-emphasis, and a spectrogram display range up to 5500 Hz. The same formant-extraction settings were applied to all speakers to maintain methodological consistency. The selected ceiling is sufficiently high to capture the first three formants for both male and female speakers in the vowel categories examined. Formant measurements were taken from the offset of frication noise of the first /f/ to the end of vowel periodicity, immediately before the onset of the following /f/. Formant values were extracted at the three time points, namely 25%, 50%, and 75%, whereas vowel durations were manually annotated by marking the onset and offset of each vowel token. Seven acoustic measures were analyzed: (i) midpoint values of F1, F2, and F3; (ii) formant dynamics for F1, F2, and F3, operationalized as the difference between measurements at the 25% and 75% time points of the vowel; and (iii) vowel duration. Both repetitions of each target vowel were included in the acoustic measurements and subsequent statistical and machine learning analyses. The 25%–75% difference measure was used to capture within-vowel spectral change while avoiding boundary regions, which are more likely to be influenced by adjacent consonants. This approach is consistent with previous work treating formant movement and spectral change across the vowel as informative dimensions of vowel realization and categorization (Chládková & Hamann, 2011; Georgiou & Paphiti, 2026).

Formant and formant-dynamic values were normalized using the Lobanov (1971) procedure (z-score normalization) to reduce inter-speaker variability attributable to anatomical differences in vocal-tract configuration (e.g., jaw/vocal-tract size) and sex/gender-related effects. For duration, the raw values were used.

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

Data analysis

Statistical tests

The acoustic differences between Greek and Albanian speakers were investigated using mixed-effects models in R (R Core Team, 2026) through the lmerTest package. Each of the seven acoustic measures (i.e., F1, F2, F3, ΔF1, ΔF2, ΔF3, and duration) served as the dependent variables, while vowel, language, and their interaction were modeled as fixed factors. Participant was modeled as a random factor. For these analyses, the vowel factor was coded using only the five vowel categories common to both Greek and Albanian (/i, e, a, o, u/). The Albanian-specific vowels /ə/ and /y/ were excluded from the mixed-effects models. To examine differences between each common vowel in Greek and Albanian, subsequent post hoc tests with the Tukey method were used.

Machine learning modeling

To simulate how Albanian vowels would be categorized using Greek vowel categories, we trained a supervised classifier on Greek productions and then applied it to Albanian tokens. Specifically, we used a Random Forest model implemented in R using the randomForest and caret packages. The training set consisted of Greek vowel tokens labeled with the five Standard Modern Greek vowel categories (/i, e, a, o, u/). Acoustic predictors were the midpoint formant measures (F1, F2, F3), formant-dynamic measures (ΔF1, ΔF2, ΔF3), and vowel duration. The vowel label was treated as a categorical outcome. Model tuning was conducted using five-fold cross-validation at the participant level, such that all tokens from a given Greek speaker were assigned to the same fold and no speaker contributed tokens to both the training and validation portions of a fold. The number of variables considered at each split (mtry) was evaluated over values 2–7, with model selection based on classification accuracy. The final Random Forest model was then fitted to all Greek data using the optimal mtry, 300 trees (ntree = 300), and a minimum terminal node size of 14. The model was subsequently applied to the Albanian dataset, using the same acoustic predictors, to assign each Albanian vowel token to one of the five Greek vowel categories. Cross-language categorization patterns were first calculated separately for each Albanian participant as row-wise percentages for each Albanian vowel and were then averaged across participants, ensuring that each speaker contributed equally to the final classification percentages. The resulting aggregated classification matrix therefore provides an estimate of how Albanian vowels map onto the Greek vowel system while accounting for speaker-level variation. To obtain a speaker-independent estimate of performance, classification accuracy was evaluated using five-fold participant-level cross-validation, with all tokens from each speaker kept within the same fold. The model achieved a mean speaker-averaged cross-validated vowel-classification accuracy of 0.92, indicating strong generalization to vowel tokens from speakers not represented in the corresponding training folds. Accuracy was also high across individual vowel categories, reaching 0.95 for /i/, 0.91 for /e/, 0.98 for /a/, 0.83 for /o/, and 0.93 for /u/.

As an additional analysis, we conducted the reverse cross-language classification to examine whether the acoustic mapping between the two vowel systems was direction-dependent. A Random Forest classifier was trained on the Albanian vowel

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

productions and then applied to the Greek vowel tokens. The reverse classifier used the same acoustic predictors, participant-level cross-validation procedure, Random Forest settings, and speaker-balanced aggregation procedure as the primary classification analysis. Because the Albanian vowel inventory contains seven vowels, whereas the Greek inventory contains five, the reverse analysis was treated as a complementary cross-language mapping rather than as a strictly symmetrical replication of the primary analysis. Participant-level cross-validation of the Albanian-trained classifier showed generally high speaker-averaged vowel-classification accuracy across the seven Albanian categories (0.81). Mean accuracy was 0.93 for /i/, 0.93 for /e/, 0.98 for /a/, 0.85 for /o/, 0.85 for /u/, 0.79 for /ə/, and 0.88 for /y/.

To examine whether Greek and Albanian productions of the shared vowel inventory could be discriminated acoustically, we trained a supervised classifier to predict language group (Greek vs. Albanian) from vowel-acoustic measures. Analyses were conducted in R using the caret, randomForest, and pROC packages. The pooled dataset contained vowel tokens from both languages, with language coded as a two-level factor (Greek, Albanian). Acoustic predictors were F1, F2, F3, ΔF1, ΔF2, ΔF3, and vowel duration. To obtain a speaker-independent evaluation, the data were split at the participant rather than the token level. Within each language group, 14 participants were randomly assigned to the training set and the remaining 7 to a held-out test set (approximately 67%/33%), ensuring that no participant contributed tokens to both datasets. Within the training set, Random Forest hyperparameters were tuned using five-fold participant-level cross-validation. Greek and Albanian participants were assigned to folds separately to maintain representation of both language groups, and all tokens from a given participant remained within the same fold. The number of variables considered at each split (mtry) was evaluated over values ranging from 2 to 7, with model selection based on cross-validated ROC performance. The Random Forest comprised 300 trees and used a minimum terminal node size of 14. The selected model was then applied to the completely unseen participants in the held-out test set. Predicted probabilities were first obtained for individual vowel tokens and then averaged within each participant to yield a single speaker-level probability of membership in the Albanian class. Participants with a mean predicted probability of at least 0.50 were classified as Albanian, whereas those with probabilities below 0.50 were classified as Greek. Speaker-level performance was quantified using accuracy, precision, recall, F1-score, and area under the Receiver Operating Characteristic (ROC) curve (AUC), with Albanian specified as the positive class for precision, recall, and F1-score.

Finally, to interpret the acoustic cues contributing most strongly to the Random Forest's language decisions, we computed SHAP values for the predicted probability of the Albanian class using the fastshap package, which approximates Shapley values via Monte Carlo sampling. A custom prediction wrapper was defined to return the model's predicted probability for the Albanian class. A background dataset comprising up to 500 randomly sampled observations from the training data was used to estimate SHAP values for observations from the held-out test participants, with 200 Monte Carlo permutations and bias adjustment applied to improve approximation accuracy. To obtain speaker-level feature contributions, the signed SHAP values for

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

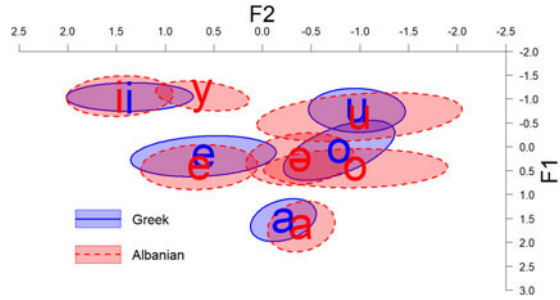


Figure 1. Normalized F1 × F2 of Greek and Albanian vowels as produced by native speakers of these languages.

each predictor were first averaged across all vowel observations produced by each participant. Global feature importance was then calculated as the mean absolute value of these participant-level SHAP contributions across speakers, ensuring that each participant contributed equally to the analysis.

Results

Acoustic differences between Greek and Albanian vowels

The descriptive results of the acoustic analyses suggest that Greek vowels /i, e, a, o, u/ occupy a largely comparable vowel space to their Albanian counterparts in terms of F1 and F2 (see Figure 1). Nevertheless, several cross-linguistic differences are apparent. Greek /o/ is realized with a more closed articulation than Albanian /o/, while Greek /a/ shows a more front realization than Albanian /a/. Additional language-specific patterns emerge in F3 and in dynamic measures. For F3, Greek /i, e, a/ exhibit lower values than the corresponding Albanian vowels, whereas Greek /o, u/ show higher F3 values than Albanian /o, u/. Dynamic measures also indicate systematic differences between the two languages. Regarding Albanian-specific categories, /ə/, absent from Greek, patterns near Greek /o/ in the F1–F2 space, and Albanian /y/ falls close to Greek /i/. Finally, vowel duration differs consistently across languages: Albanian tokens are shorter than Greek tokens for all shared vowels (see Figure 2). Table 1 reports the unnormalized (raw) values for each acoustic measure for Greek and Albanian vowels.

The mixed-effects models revealed that, within Albanian, several vowel categories differed significantly from the reference vowel /a/, particularly for F1, F2, F3, and duration. Language-related effects were more limited, as significant differences between Greek and Albanian for the reference vowel /a/ were observed only for F2, ΔF3, and duration. Several significant interaction effects were also identified, indicating that Greek–Albanian differences were vowel-specific rather than uniform across all vowel categories. The full results of the mixed-effects models are reported in Table 2. The post hoc comparisons for vowels shared by the two languages showed selective vowel-specific differences. Specifically, significant differences were found for all vowels except /i/ in F1; for vowel /a/ in F2; for vowels /i, e/ in F3; for vowel /u/ in ΔF1; for vowel /o/ in ΔF2; for vowels /o, u/ in ΔF3; and for all vowels in duration. These results are presented in Table 3.

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

Table 1. Unnormalized values of Greek and Albanian vowels for each acoustic measure

Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)	ΔF1 (Hz)	ΔF2 (Hz)	ΔF3 (Hz)	Duration (ms)
Greek							
i	383 (47)	2266 (410)	2707 (328)	−22 (55)	−24 (290)	41 (323)	131 (32)
e	589 (98)	1790 (361)	2555 (309)	−26 (67)	25 (245)	66 (302)	149 (35)
a	810 (147)	1310 (154)	2575 (340)	−9 (77)	5 (86)	0 (320)	162 (35)
o	562 (90)	954 (235)	2764 (296)	14 (64)	62 (292)	36 (226)	152 (32)
u	425 (53)	850 (188)	2749 (293)	−26 (102)	61 (553)	−25 (432)	129 (35)
Albanian							
i	368 (51)	2286 (291)	2851 (286)	−7 (42)	−30 (178)	−85 (149)	106 (33)
e	569 (76)	1852 (286)	2666 (306)	−16 (48)	−13 (195)	7 (156)	112 (29)
a	737 (107)	1285 (183)	2564 (264)	−33 (93)	−40 (101)	30 (143)	122 (30)
o	570 (68)	997 (339)	2687 (300)	−49 (81)	−147 (342)	−54 (175)	116 (29)
u	422 (41)	939 (375)	2695 (282)	−13 (75)	164 (436)	151 (401)	94 (29)
ə	369 (42)	1817 (183)	2472 (248)	−16 (41)	−29 (162)	43 (142)	107 (33)
y	545 (74)	1284 (219)	2600 (270)	−26 (71)	−45 (132)	−11 (113)	103 (30)

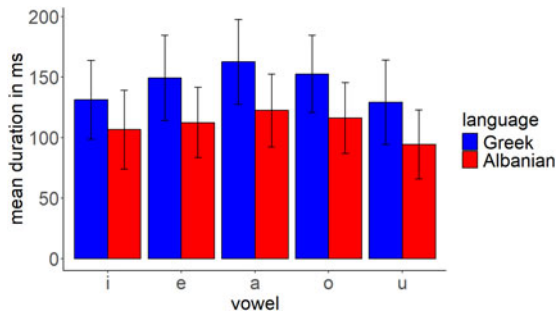


Figure 2. Durations of Greek and Albanian common vowels as produced by native speakers of these languages.

Cross-language classification of Greek and Albanian vowels

The results of the classification simulation indicated a largely one-to-one correspondence between Albanian vowels and Greek vowel categories for the vowels shared by the two languages. Albanian /i/ was classified almost exclusively as Greek /i/. Similarly, Albanian /e/ was predominantly classified as Greek /e/. Albanian /a/ was fully classified as Greek /a/. Albanian /o/ and /u/ also showed high classification accuracy, with Albanian /o/ most frequently mapped onto Greek /o/, while Albanian /u/ was mainly classified as Greek /u/, with some below-chance overlap with Greek /o/. For vowels that are not part of the Greek vowel inventory, the model assigned them to the closest Greek categories. Albanian /ə/ was most often classified as Greek

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

Table 2. Results of the mixed-effects models for each acoustic measure. The intercept terms are vowel /a/ and language Albanian

	Estimate	Std. Error	*t*–value	*p*–value
F1	1.67	0.05	35.94	<0.001
(Intercept)	−1.24	0.07	−18.93	<0.001
vowele	−2.73	0.07	−41.58	<0.001
voweli	−1.22	0.07	−18.32	<0.001
vowelo	−2.29	0.06	−35.31	<0.001
vowelu	−0.13	0.07	−2.05	0.04
languageGreek	−0.09	0.09	−1.01	0.31
vowele:languageGreek	0.15	0.09	1.65	0.10
voweli:languageGreek	−0.24	0.09	−2.58	0.01
vowelo:languageGreek	0.00	0.09	−0.05	0.96
vowelu:languageGreek	1.67	0.05	35.94	<0.001
F2				
(Intercept)	−0.40	0.07	−6.00	<0.01
vowele	1.05	0.09	11.18	<0.001
voweli	1.86	0.09	19.89	<0.001
vowelo	−0.56	0.09	−5.88	<0.01
vowelu	−0.60	0.09	−6.47	<0.01
languageGreek	0.18	0.09	1.95	0.05
vowele:languageGreek	−0.23	0.13	−1.71	0.09
voweli:languageGreek	−0.29	0.13	−2.16	0.03
vowelo:languageGreek	−0.01	0.13	−0.11	0.91
vowelu:languageGreek	−0.16	0.13	−1.18	0.24
F3				
(Intercept)	−0.38	0.14	−2.73	0.01
vowele	0.59	0.20	3.00	<0.01
voweli	1.34	0.20	6.78	<0.01
vowelo	0.44	0.20	2.21	0.03
vowelu	0.46	0.19	2.35	0.02
languageGreek	0.07	0.20	0.34	0.73
vowele:languageGreek	−0.74	0.28	−2.64	0.01
voweli:languageGreek	−1.00	0.28	−3.58	<0.01
vowelo:languageGreek	0.31	0.28	1.11	0.27
vowelu:languageGreek	0.16	0.28	0.56	0.57

(*Continued*)

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

Table 2. (*Continued*)

	Estimate	Std. Error	*t*−value	*p*−value
ΔF1				
(Intercept)	0.02	0.14	0.13	0.90
vowele	−0.01	0.19	−0.06	0.95
voweli	−0.39	0.19	−2.00	0.05
vowelo	−0.22	0.20	−1.12	0.26
vowelu	0.48	0.19	2.51	0.01
languageGreek	−0.04	0.19	−0.21	0.84
vowele:languageGreek	0.10	0.27	0.37	0.71
voweli:languageGreek	0.41	0.27	1.50	0.14
vowelo:languageGreek	0.35	0.28	1.25	0.21
vowelu:languageGreek	−0.61	0.27	−2.24	0.03
ΔF2				
(Intercept)	−0.09	0.06	−1.35	0.18
vowele	0.14	0.09	1.49	0.14
voweli	0.19	0.09	2.04	0.04
vowelo	−0.12	0.09	−1.25	0.21
vowelu	0.20	0.09	2.21	0.03
languageGreek	0.04	0.09	0.48	0.63
vowele:languageGreek	−0.18	0.13	−1.39	0.17
voweli:languageGreek	−0.15	0.13	−1.19	0.24
vowelo:languageGreek	0.31	0.13	2.39	0.02
vowelu:languageGreek	−0.17	0.13	−1.29	0.20
ΔF3				
(Intercept)				
vowele	−0.05	0.08	−0.58	0.56
voweli	0.07	0.11	0.65	0.52
vowelo	0.05	0.11	0.44	0.66
vowelu	−0.20	0.11	−1.81	0.07
languageGreek	0.36	0.11	3.34	<0.01
vowele:languageGreek	0.01	0.11	0.09	0.93
voweli:languageGreek	−0.06	0.16	−0.38	0.70
vowelo:languageGreek	−0.07	0.16	−0.48	0.63
vowelu:languageGreek	0.31	0.16	1.94	0.05

(*Continued*)

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

Table 2. (*Continued*)

	Estimate	Std. Error	*t*−value	*p*−value
Duration				
(Intercept)	122.24	6.54	18.71	<0.001
vowele	−9.93	3.73	−2.66	0.01
voweli	−15.80	3.73	−4.23	<0.01
vowelo	−6.88	3.79	−1.82	0.07
vowelu	−27.43	3.70	−7.42	<0.01
languageGreek	40.18	9.24	4.35	<0.01
vowele:languageGreek	−3.30	5.28	−0.63	0.53
voweli:languageGreek	−15.57	5.28	−2.95	<0.01
vowelo:languageGreek	−3.13	5.32	−0.59	0.56
vowelu:languageGreek	−5.84	5.26	−1.11	0.27

/o/, followed by Greek /e/, suggesting that /ə/ occupies the mid region of the Greek vowel space, with a stronger tendency toward the back vowel /o/. Albanian /y/ was classified primarily as Greek /i/, with the remaining tokens mainly mapped to Greek /u/, indicating an alignment with the high region of the Greek vowel space. Overall, classification errors for shared vowels were limited and systematic, while Albanian-specific vowels showed predictable assimilation to the nearest Greek vowel categories based on their acoustic properties. These classification patterns are summarized in Table 4. A multivariate Euclidean distance analysis based on all seven acoustic measures corroborates these results by showing that Albanian /ə/ was acoustically closest to Greek /o/ (distance = 1.67, 95% CI [1.32, 2.04]), followed very closely by Greek /e/ (distance = 1.74, 95% CI [1.51, 1.99]) Albanian /y/ was most similar to Greek /i/ (distance = 1.44, 95% CI [1.16, 1.75]), with substantially greater distances to the other vowels.

To complement the primary Albanian-to-Greek classification, we additionally examined the reverse mapping by training the classifier on Albanian vowel productions and applying it to the Greek vowels. The analysis showed a largely systematic mapping of the Greek vowels onto Albanian vowel categories. Greek /i, e, a/ were predominantly classified as their corresponding Albanian vowel categories, indicating substantial acoustic similarity between the shared vowels in these regions of the vowel space. Greek /u/ also showed a strong correspondence with Albanian /u/. In contrast, Greek /o/ exhibited a more dispersed classification pattern, with productions distributed across several Albanian categories, including /o, u, ə/. Greek /i/ also showed a secondary tendency to be classified as Albanian /y/. Table 5 presents the results of this classification.

Classification of speakers based on common Greek and Albanian vowels

Table 6 summarizes the speaker-level performance of the machine-learning classifier in distinguishing Greek and Albanian speakers based on the vowels shared by

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

Table 3. Results of the post hoc tests for the common vowels between Greek and Albanian

Greek vs Albanian	Estimate	Std. Error	*t*−value	*p*−value
F1				
i	−0.02	0.07	−0.28	0.78
e	0.23	0.07	3.48	<0.01
a	0.13	0.07	2.05	0.04
o	0.38	0.07	5.65	<.0001
u	0.14	0.06	2.14	0.03
F2				
i	0.10	0.09	1.10	0.27
e	0.04	0.09	0.46	0.64
a	−0.18	0.09	−1.95	0.05
o	−0.17	0.09	−1.77	0.08
u	−0.03	0.09	−0.30	0.77
F3				
i	0.93	0.20	4.72	<.0001
e	0.67	0.20	3.39	<0.01
a	−0.07	0.20	−0.34	0.73
o	−0.38	0.20	−1.90	0.06
u	−0.22	0.20	−1.15	0.25
ΔF1				
i	−0.37	0.19	−1.91	0.06
e	−0.06	0.19	−0.31	0.75
a	0.04	0.19	0.21	0.84
o	−0.31	0.20	−1.56	0.12
u	0.65	0.19	3.40	<0.01
ΔF2				
i	0.11	0.09	1.20	0.23
e	0.13	0.09	1.48	0.14
a	−0.04	0.09	−0.48	0.63
o	−0.35	0.09	−3.84	<0.01
u	0.12	0.09	1.36	0.18
ΔF3				
i	0.06	0.11	0.58	0.56
e	0.05	0.11	0.45	0.66

(*Continued*)

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

Table 3. (*Continued*)

Greek vs Albanian	Estimate	Std. Error	*t*–value	*p*–value
a	−0.01	0.11	−0.09	0.93
o	−0.32	0.11	−2.82	0.01
u	0.27	0.11	2.44	0.02
Duration				
i	−24.60	9.24	−2.66	0.01
e	−36.90	9.24	−3.99	<0.01
a	−40.20	9.24	−4.35	<0.01
o	−37.10	9.26	−4.00	<0.01
u	−34.30	9.23	−3.72	<0.01

Table 4. Simulation of the classification of Albanian vowels in terms of Greek vowel categories. Blue cells show the dominant classification, and bold numbers show classifications above chance (i.e., ≥20%)

		Greek				
vowel		i	e	a	o	u
Albanian	i	**98**	0	0	0	2
	e	2	**91**	2	5	0
	a	0	0	**100**	0	0
	o	3	0	3	**94**	0
	u	2	2	0	11	**85**
	ə	0	**26**	0	**67**	7
	y	**64**	5	0	0	**31**

Table 5. Simulation of the classification of Greek vowels in terms of Albanian vowel categories. Blue cells show the dominant classification, and bold numbers show classifications above chance (i.e., ≥14%)

		Albanian						
vowel		i	e	a	o	u	ə	y
Greek	i	**84**	0	0	0	2	0	**14**
	e	0	**88**	0	5	0	7	0
	a	0	0	**98**	2	0	0	0
	o	0	2	2	**53**	**17**	**26**	0
	u	0	0	0	7	**93**	0	0

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

Table 6. Evaluation performance of the machine learning classifier

Metric	Score
Accuracy	0.86
Precision	1.00
Recall	0.71
F1-score	0.83
AUC	0.96

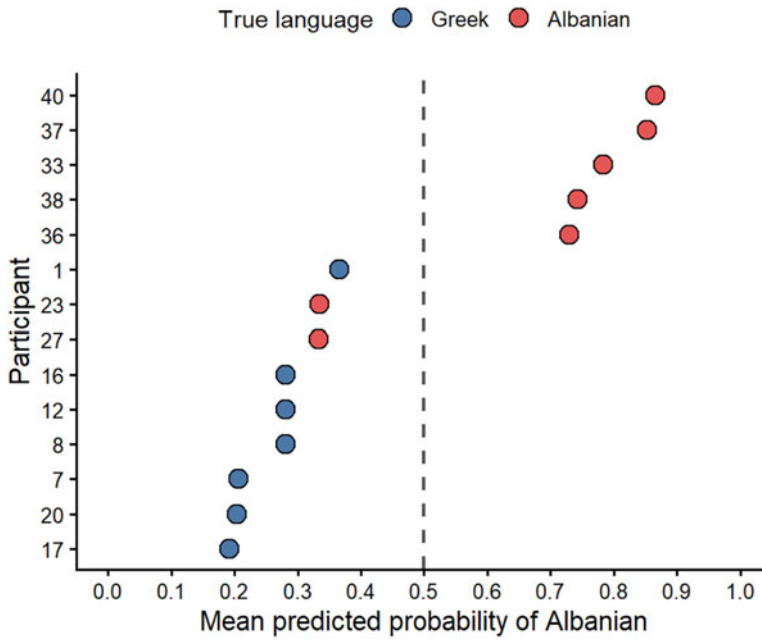


Figure 3. Speaker-level predicted probability of Albanian for unseen test participants. Each point represents one participant, with probabilities averaged across that participant's vowel tokens.

the two languages. The classifier achieved an overall accuracy of 0.86, indicating a high proportion of correctly classified participants by language. Precision was 1.00, showing that all speakers classified as Albanian were indeed Albanian, whereas recall was 0.71, indicating that 71% of the Albanian speakers were correctly identified. The resulting F1-score of 0.83 reflects this balance between precision and recall. Finally, the AUC of 0.96 indicated strong discriminative ability between Greek and Albanian speakers.

Figure 3 displays the speaker-level predictions of the Random Forest language classification model for the unseen test participants. Each point represents one participant and shows the mean predicted probability of being Albanian, averaged across that participant's vowel tokens. The dashed vertical line at 0.50 indicates the classification threshold, with probabilities below 0.50 classified as Greek and

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

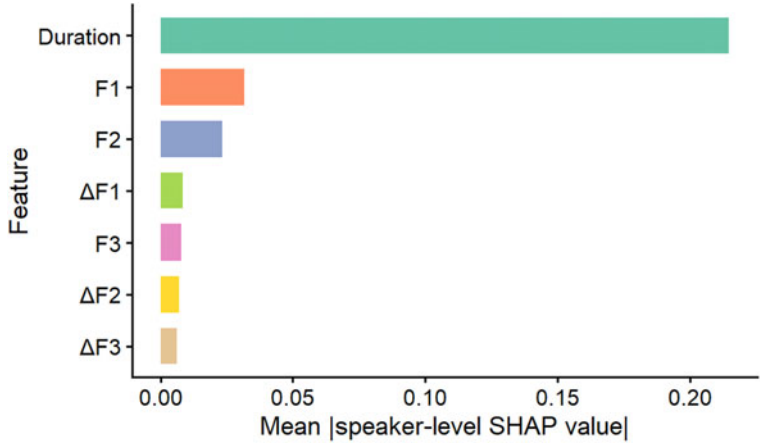


Figure 4. Speaker-level mean absolute SHAP values for each acoustic feature used in the language classification model.

probabilities at or above 0.50 classified as Albanian. All Greek participants were correctly classified, with predicted probabilities remaining below the decision threshold. Five of the seven Albanian participants were also correctly classified, whereas two Albanian participants received probabilities below 0.50 and were consequently classified as Greek.

Feature importance analysis

The speaker-level SHAP feature importance analysis (Figure 4) showed that duration was by far the most influential predictor in the classification model, with a substantially larger mean absolute SHAP value than all other features. Among the spectral measures, F1 was the most important predictor, followed by F2, whereas F3 made a comparatively smaller contribution. The dynamic formant measures had relatively limited influence on the model, with ΔF1 showing the largest contribution among the dynamic measures, followed by ΔF2 and ΔF3. Overall, these results indicate that the classifier relied predominantly on temporal information, particularly vowel duration, with static spectral properties—especially F1 and F2—providing additional discriminatory information. Dynamic formant measures contributed comparatively little to the speaker-level distinction between Greek and Albanian.

Discussion

The present study treated Greek–Albanian vowel similarity as a multi-dimensional problem and asked three explicit questions: (1) how similar Greek and Albanian realizations of the *shared* vowels are when static, dynamic, and temporal cues are considered together; (2) how vowels would be categorized across the Greek and Albanian vowel systems, with Albanian-to-Greek classification as the primary mapping and Greek-to-Albanian classification as a complementary reverse analysis (a computational analogue of cross-language category mapping); and (3) which acoustic cues most strongly drive cross-language discriminability and model

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

decisions, as revealed by Random Forest feature attribution with SHAP. Across analyses, the results converge on a consistent interpretation: Greek and Albanian show substantial overlap in canonical midpoint vowel-space representations, but they differ systematically in timing and in a smaller set of spectral and dynamic properties. These differences are sufficient to produce robust computational discriminability and highly interpretable mapping patterns that align with similarity-based accounts of categorization.

Addressing the first research question, the acoustic analyses confirm that the shared five-vowel inventory /i, e, a, o, u/ is broadly comparable across languages in its overall spatial organization, which is consistent with long-standing evidence that vowel quality is strongly captured by formant structure and that many languages converge on dispersed solutions under articulatory–acoustic constraints (Hillenbrand et al., 1995; Liljencrants & Lindblom, 1972; Maddieson, 1984; Peterson & Barney, 1952). The present dataset nevertheless shows that shared inventory does not entail identical implementation, echoing typological and cross-language perception work showing that phonological overlap can coexist with substantial phonetic non-equivalence (e.g., see Cho, 2025; Millet et al., 2025). In the statistical analysis, vowel category produced strong effects across measures (as expected given the systematic acoustic structure of vowel systems), while the analysis revealed limited language effects at the level of the reference vowel. Crucially, the post hoc tests indicated that cross-language differences are selective but patterned: for midpoint F1, differences were observed for all shared vowels except/i/; for midpoint F2, a reliable difference emerged specifically for /a/; for midpoint F3, differences were concentrated on /i/ and /e/; for dynamics, differences were most clearly expressed for /u/ in ΔF1, for /o/ in ΔF2, and for /o/ and /u/ in ΔF3; and for duration, differences were robust across all shared vowels. Substantively, the pattern for midpoint measures supports a "mostly aligned, selectively shifted" conclusion. Greek /o/ is described as more closed than Albanian /o/, and Greek /a/ is more front than Albanian /a/. These are the kinds of implementation differences expected when a seven-vowel system (Albanian) must maintain contrast not only among /i, e, a, o, u/ but also with /ə/ and /y/, potentially reshaping dispersion and crowding pressures across the system (Al-Tamimi & Ferragne, 2005; Schwartz et al., 1997). Contemporary descriptive work on Albanian confirms a seven-vowel system that includes /ə/ and /y/ and documents systematic vowel-space organization (Coretta et al., 2023), providing a concrete phonetic basis for expecting system-wide interactions rather than purely category-local shifts. Importantly, these midpoint differences were not uniform across all vowels, which is consistent with the idea that cross-language similarity is structured unevenly, with certain categories remaining near-equivalent while others drift, creating predictable "pressure points" for mapping and discriminability.

A key contribution of the present study is that it demonstrates why focusing on midpoint F1–F2 alone can understate cross-language distance. Across the shared vowels, Albanian productions were consistently shorter than Greek productions, and duration differences were reported as significant across vowels in the post hoc comparisons. This is theoretically and methodologically consequential, since temporal properties often provide crucial secondary cues for category identification and contrast maintenance (Hillenbrand et al., 1995; Peterson & Barney, 1952), and duration can also act as a stable language-or variety-level signature in controlled

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

speech styles. The fact that duration differences were pervasive suggests that Greek–Albanian similarity, in this dataset, is not simply a matter of "where vowels sit in the plane," but also of "how long the system gives listeners to sample the signal," which can shape both perceptual assimilation and machine classification when temporal information is available. The prioritization of the duration cue corroborates the results of previous studies, particularly Cebrian (2006), who showed that L2 listeners can overrely on duration even when vowel length is not phonemic in their L1, and Mora and Fullana (2007), who found that advanced learners of English often weight durational differences more than formant frequencies in production.

The results for F3 and dynamics further reinforce a multi-dimensional view. The language differences in F3 (especially for /i/ and /e/) suggest language-specific spectral shaping beyond the primary F1–F2 dimensions. While F3 is often less central to canonical vowel-space plots, it can reflect articulatory setting and rounding-related spectral consequences and can contribute to category differentiation in subtle ways (Hillenbrand et al., 1995). The dynamic measures (ΔF1–ΔF3) showed fewer and more vowel-specific differences than duration and midpoint measures, but the significant contrasts for /o/ and /u/ are conceptually meaningful because dynamic information can carry language-specific structure and can be exploited by listeners, particularly when midpoint overlap is high or when coarticulation and vowel-inherent spectral change are important (e.g., Nittrouer, 2007; Williams et al., 2018). Recent work modeling formant trajectories as multidimensional time series in L2 English further underscores that trajectory-sensitive representations can reveal structure not captured by midpoints alone (even if the magnitude of those effects depends on context and measurement choices), supporting the broader motivation for including dynamics in typological and cross-language analyses (Adem, 2026).

The primary analysis addressing the second research question asked how Albanian vowels would be categorized if "heard" through Greek vowel categories, operationalized by training a supervised classifier on Greek productions and applying it to Albanian tokens (for a similar approach, see Escudero et al., 2012). The mapping results showed a near one-to-one correspondence for the vowels shared by the two languages. More specifically, Albanian /i/ was classified almost exclusively as Greek /i/, Albanian /e/ predominantly as Greek /e/, Albanian /a/ categorically as Greek /a/, and Albanian /o/ mainly as Greek /o/, with only limited and structured confusions in Albanian /u/, which was occasionally classified as Greek /o/. This is precisely the mapping profile one would expect under similarity-driven category alignment: when categories are sufficiently separated in the relevant cue space, mapping is stable; when categories are closer (often in the back region, where rounding and spacing can create overlap), confusions increase (Best & Tyler, 2007; Flege & Bohn, 2021; Georgiou, 2021). The vowel-specific acoustic findings support this interpretation: /u/ shows a dynamic difference (ΔF1) and participates in ΔF3 contrasts, all of which can contribute to a mild /u/ ambiguity even when the overall vowel spaces are similar. Still, the most important point is that for the vowels the two languages share, the model mostly gives a same-to-same mapping. This means that, in terms of the acoustic cues available to the classifier, the shared vowel categories are close enough across the two languages that they are treated as functionally equivalent most of the time, even though the statistical analyses show selective phonetic differences in some measures. This pattern is consistent with

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

dispersion-based accounts of vowel systems, which propose that vowel categories are shaped by system-wide contrast relations and that differences in inventory structure may influence their phonetic implementation (Al-Tamimi & Ferragne, 2005; Recasens & Espinosa, 2009; Schwartz et al., 1997).

The Albanian-only categories are especially informative because they show how a richer vowel inventory is systematically mapped onto a smaller one. This is particularly relevant given that much prior work has focused on the reverse problem, namely classifying or assimilating vowels that are absent in the listener's language when that listener's vowel system is comparatively less dense (Baigorri et al., 2019; Lengeris, 2009). Albanian /ə/ was most frequently classified as Greek /o/ (with substantial secondary mapping to /e/), indicating that /ə/ occupies an intermediate region that Greek must interpret using its available mid/back categories. Albanian /y/ was classified primarily as Greek /i/ with a smaller proportion mapped to Greek /u/, which is the expected outcome when Greek lacks a front rounded category and classification is driven by competing cues, namely, frontness draws /y/ toward /i/, while rounding can pull some tokens toward /u/. This pattern aligns with the logic of PAM/PAM-L2, in which nonnative phones are assimilated to the closest native categories, while intermediate acoustic positioning or cue conflict may lead to weaker, less categorical mappings distributed across more than one native category rather than clean single-category alignment (Best & Tyler, 2007). It also generates clear, testable perceptual predictions based on UPM where Greek listeners should show high accuracy for Albanian /i, e, a, o, u/ but lower for /ə, y/, meaning that contrasts such as /i – y/ and /ə – o/ would be difficult, because they create partial overlap (i.e., both L2 vowels share at least one above-chance L1 category, respectively). Finally, it is consistent with SLM-r, which predicts that nonnative sounds may be assimilated to acoustically similar L1 categories when they are close to existing native categories (Flege & Bohn, 2021). It is important to note that although the classifier provides a useful computational model of category mapping based on acoustic similarity, the cue-weighting strategies learned by machine learning algorithms may differ from those employed by human listeners. Consequently, the observed assimilation patterns should not be interpreted as direct evidence of human perceptual behavior. Rather, they represent predictions derived from the acoustic structure of the data.

The complementary reverse classification further clarified the directionality of these cross-language relationships. When the classifier was trained on the larger Albanian vowel inventory and applied to Greek productions, Greek/ i, e, a, u/ were predominantly assigned to their corresponding Albanian categories, reinforcing the substantial acoustic correspondence observed for the shared vowels in the primary analysis. However, the reverse mapping was not entirely symmetrical: Greek /o/ showed a more distributed mapping across Albanian /o/, /u/, and /ə/, while Greek /i/ showed some secondary association with Albanian /y/. These patterns suggest that cross-language category correspondence depends partly on the structure of the vowel inventory used as the reference system. In particular, the additional Albanian categories provide alternative classification targets that are unavailable when Albanian vowels are mapped onto the smaller five-vowel Greek system. More broadly, the directional differences observed here are consistent with evidence that cross-language vowel similarity is not necessarily reciprocal; the mapping of a vowel

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

from Language A onto Language B may differ from the corresponding mapping from Language B onto Language A because similarity is evaluated relative to the category structure and cue-weighting patterns of the reference language (Cebrian, 2021, 2022; Flege et al., 1994).

The present mapping results provide a possible phonetic basis for the patterns previously reported in Albanian speakers' acquisition and use of Greek vowels (Georgiou & Giannakou, 2024; Kaskampa & Georgiou, 2025). In the present study, the largely one-to-one mappings of the shared vowels indicate strong cross-language category correspondence, whereas the acoustic analyses show that these corresponding categories are not fully equivalent phonetically. This combination is particularly relevant to L2 speech learning, as Albanian speakers may establish broad correspondence between their existing /i, e, a, o, u/ categories and the respective Greek vowels while nevertheless retaining L1-related phonetic properties in their Greek productions. In previous work, the residual phonetic differences documented in Albanian L2 and heritage speakers of Greek may reflect, at least partly, differences already present between the corresponding native Greek and Albanian categories identified here. In contrast, the present mappings of Albanian /ə/ and /y/ suggest a different potential learning situation because these vowels lack direct phonological counterparts in Greek. Therefore, their distributed mappings generate the prediction that these categories may present greater cross-language categorization difficulty than the shared vowels. However, given the limited direct perceptual evidence for these specific Greek–Albanian vowel relationships, this interpretation should be regarded as a testable prediction rather than as evidence of perceptual mapping.

The third research question concerned model-based cue weighting and discriminability, aiming to investigate which acoustic dimensions most strongly differentiate Greek from Albanian vowel realizations in the Random Forest language classifier, and what SHAP reveals about the model's reliance on cues. The classifier achieved strong held-out performance within the present dataset, and the ROC analysis suggested that the acoustic measures contained systematic cross-language information. The SHAP analysis makes the classifier's decision process explicit, showing that duration was by far the most influential predictor, followed by F1 and F2. F3 had a smaller contribution, while the dynamic measures (ΔF1–ΔF3) contributed comparatively little to the model's speaker-level predictions. This finding has interpretive relevance beyond model performance alone. First, it validates the multi-cue motivation of the study, as the strongest cross-language separator in this dataset is temporal structure, and without duration, the apparent similarity in midpoint F1–F2 would likely lead to substantially weaker discriminability. Second, it ties the computational results back to phonetic interpretation: SHAP is not merely ranking features; it is quantifying how the model's language decisions are built from the same dimensions that mixed-effects models identify as reliably different (especially duration, and selectively F2 and dynamic measures). Methodologically, this illustrates that XAI is increasingly advocated in speech and acoustics applications since it enables researchers to move from "the model works" to "the model works because it is exploiting these specific acoustic regularities," a shift emphasized in recent reviews of explainability in acoustics and applied audio classification (McCarthy et al., 2025). Importantly, these SHAP values reflect the cue-weighting

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

behavior of the classifier rather than that of human listeners. Consequently, the results should be interpreted as model-based estimates of acoustic cue importance.

The prominence of duration invites careful interpretation and is particularly relevant in light of previous findings on Greek–Albanian bilingual speech. Georgiou and Giannakou (2024) reported durational differences between Albanian heritage speakers and monolingual Greek speakers across the Greek vowel inventory. The systematic temporal differences observed between native Greek and Albanian vowels in the present study therefore identify a plausible L1 phonetic dimension along which cross-language influence may persist when Albanian-background speakers acquire or use Greek. More broadly, the convergence between the native-language comparison, the classifier's strong reliance on duration, and previously documented durational differences in Albanian-background Greek speech suggests that temporal structure may be especially important for understanding Greek–Albanian L2 phonetic interaction. Duration can contribute to vowel contrast and identification alongside spectral information (Hillenbrand et al., 1995; Peterson & Barney, 1952), while also reflecting broader language-specific patterns of timing and prosodic organization. Because the present study employed different carrier phrases in the two languages within a reading task, the observed durational differences may capture not only vowel-intrinsic timing but also variation associated with phrase-level rhythm, speech rate, and language-specific prosodic implementation in citation-like speech. Although the controlled /fVf/ frame, comparable task structure, and consistent segmentation procedure reduced some sources of variability, the design cannot fully disentangle segmental duration from these broader temporal influences. This interpretive caution is particularly relevant given the classifier's strong dependence on duration, since the measurement of temporal features necessarily depends on how vowel boundaries are defined. Recent work on vowel classification has shown that segmentation and measurement choices can influence machine-learning performance and, consequently, which acoustic features emerge as most diagnostic (Idrisoglu et al., 2025).

The comparatively low SHAP importance of dynamic measures also deserves a phonetic explanation. A large literature shows that dynamic cues and vowel-inherent spectral change (VISC) can be central to vowel identification and naturalness judgments, and recent experimental work has asked how much formant detail is required for identifying and rating vowels, underscoring that trajectory information can matter depending on task and signal properties (Jibson, 2022). In the present design, however, dynamics were operationalized as a single difference between 75% and 25% time points, and the /fVf/ context was intentionally chosen to minimize coarticulatory variability. Both choices plausibly reduce the incremental contribution of dynamics over and above midpoints and duration. That is, dynamic cues may be "real" but under-represented by a coarse Δ measure in a highly controlled frame. Recent trajectory-centric approaches model formant movement as a time series or use richer parameterizations (e.g., multidimensional trajectory measures), and such approaches have been shown to reveal dynamic structure more clearly than midpoint-plus-delta summaries in some contexts (Brandt et al., 2021). A direct methodological implication is that the present conclusion should be interpreted cautiously: in this dataset, as measured here, cross-language differentiation relies mainly on duration and static spectral cues, while dynamics provide

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

secondary support; it does not follow that dynamics are unimportant in Greek–Albanian vowel systems in general.

From a typological and areal perspective, these results shed light on patterns of phonetic stability versus divergence in vowel systems within the Balkan linguistic area, where prolonged multilingual contact has fostered convergence in grammar, prosody, and select phonological features, even across genetically distant families (Friedman & Joseph, 2017, 2025; Joseph, 2003). This selective pattern found in the present study aligns with observations that Balkan phonological convergence is often partial, localized, and domain-specific (e.g., central-vowel innovations), underscoring that areal contact tends to promote shared innovations unevenly while preserving underlying genetic and typological distinctiveness in vowel implementation (Friedman & Joseph, 2025). These typological patterns acquire added significance in contemporary bilingual contexts. The present acoustic and computational comparison holds particular importance given the substantial bilingual Greek-Albanian population, especially among Albanian-heritage communities in Greece, where large-scale migration since the 1990s has created generations of bilingual speakers who maintain varying degrees of Albanian proficiency alongside Greek dominance. Similar bilingual dynamics characterize the Greek minority in southern Albania, where functional bilingualism in both languages is common. These contexts underscore the value of establishing precise baseline phonetic differences between the two languages: such data provide an essential foundation for investigating cross-language perception, category assimilation patterns, cue-weighting shifts, and heritage-language maintenance in bilingual populations, particularly those with migrant backgrounds in Greece, where ongoing research explores these phenomena (Georgiou & Giannakou, 2024).

Conclusions

This study provided a multi-dimensional acoustic and computational analysis of vowel production in Greek and Albanian, two languages that share a core five-vowel inventory. The findings suggest that a shared phonological inventory does not necessarily entail full phonetic equivalence. While the vowels /i, e, a, o, u/ occupy largely overlapping regions in the canonical F1–F2 space, systematic differences were identified in duration, selected spectral properties, and formant dynamics. Crucially, a supervised classifier reliably distinguished between Greek and Albanian speakers, demonstrating that these subtle, multi-cue differences are acoustically systematic and discriminable. The simulations of cross-language category mapping revealed largely corresponding mappings for shared vowels in both directions, although the reciprocal mappings were not fully symmetrical. In the primary Albanian-to-Greek analysis, Albanian-specific /ə/ and /y/ were predictably mapped to the closest Greek categories (/o/ and /i/, respectively). XAI analysis identified vowel duration as the primary cue driving classifier-based discriminability, followed by static formant measures, highlighting that temporal properties can be a key dimension of cross-language phonetic divergence even when spectral spaces are aligned.

These results carry theoretical implications for cross-language speech perception and L2 speech learning, particularly in light of evidence that Albanian learners and heritage speakers of Greek may retain fine-grained phonetic differences in otherwise

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

corresponding vowel categories. The present findings provide a possible acoustic basis for such patterns by showing that substantial category-level correspondence can coexist with systematic differences in temporal and spectral dimensions. The study also underscores the methodological value of integrating supervised classification with interpretability tools like SHAP for phonetic research, moving beyond descriptive statistics to a quantitative account of cue weighting in cross-language comparison. The conclusions are tempered by certain limitations. The use of different carrier phrases may have influenced the robust duration differences, and the operationalization of formant dynamics via a single delta measure may have understated their role. Furthermore, the /fVf/ context represents a controlled speech style that may not fully reflect natural speech, and the relatively small sample size warrants cautious interpretation of the classifier's performance metrics. Future work should validate the predicted assimilation patterns with perceptual experiments, employ more nuanced trajectory analyses, and replicate the machine learning analyses with larger datasets.

Replication package. All experimental materials, analysis code, and data are available at https://doi.org/10.17605/OSF.IO/JZTEK.

Acknowledgements. This study was supported by the Phonetic Lab of the University of Nicosia. We thank all participants for taking part in this study. We also thank Sonila Mehillaj for assisting with the collection of the Albanian data. Written informed consent was obtained from all participants prior to participation.

Competing interests. We have no competing interests to disclose.

References

Adem, H. (2026). Dynamic formant trajectories in L2 English: Evidence from Arabic-speaking adolescent learners. *Journal of Phonetics*, **114**, 101471.

Al-Tamimi, J. E., & Ferragne, E. (2005, September). Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. In Interspeech (pp. 2465–2468).

Alispahic, S., Mulak, K. E., & Escudero, P. (2017). Acoustic properties predict perception of unfamiliar Dutch vowels by adult Australian English and Peruvian Spanish listeners. *Frontiers in Psychology*, **8**, 52.

Arvaniti, A. (1999). Standard Modern Greek. *Journal of the International Phonetic Association*, **29**(2), 167–172.

Baigorri, M., Campanelli, L., & Levy, E. S. (2019). Perception of American–English vowels by early and late Spanish–English bilinguals. *Language and Speech*, **62**(4), 681–700.

Benway, N. R., Preston, J. L., Salekin, A., Hitchcock, E., & McAllister, T. (2024). Evaluating acoustic representations and normalization for rhoticity classification in children with speech sound disorders. *JASA Express Letters*, **4**(2), 025201.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins Publishing Company.

Boersma, P., & Weenink, D. (2025). *Praat: Doing Phonetics by Computer [Computer software]*. https://www.praat.org/

Brandt, E., Möbius, B., & Andreeva, B. (2021). Dynamic formant trajectories in German read speech: Impact of predictability and prominence. *Frontiers in Communication*, **6**, 643528.

Breiman, L. (2001). Random forests. *Machine Learning*, **45**(1), 5–32.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, **34**(3), 372–387.

Cebrian, J. (2021). Perception of English and Catalan vowels by English and Catalan listeners: A study of reciprocal cross-linguistic similarity. *The Journal of the Acoustical Society of America*, **149**(4), 2671–2685.

Cebrian, J. (2022). Perception of English and Catalan vowels by English and Catalan listeners. Part II. Perceptual vs ecphoric similarity. *The Journal of the Acoustical Society of America*, **152**(5), 2781–2793.

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

Chládková, K., & Hamann, S. (2011, August). High Vowels in Southern British English:/u/-fronting Does Not Result in Merger. In *ICPhS* (Vol. 17, pp. 476–479).

Cho, T. (2025). Advancements of phonetics in the 21st century: Phonetic universals, language variation, and phonetic grammar. *Journal of Phonetics*, **111**, 101426.

Coretta, S., Riverin-Coutlée, J., Kapia, E., & Nichols, S. (2023). Northern Tosk Albanian. *Journal of the International Phonetic Association*, **53**(3), 1122–1144.

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., & Ranjan, R. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, **55**(9), 1–33.

Elvin, J., Williams, D., Shaw, J. A., Best, C. T., & Escudero, P. (2021). The role of acoustic similarity and non-native categorisation in predicting non-native discrimination: Brazilian Portuguese Vowels by English vs. Spanish Listeners. *Languages*, **6**(1), 44.

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics*, **40**(2), 280–288.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–25). Cambridge University Press.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *The Journal of the Acoustical Society of America*, **95**(6), 3623–3641.

Friedman, V. A., & Joseph, B. D. (2017). Reassessing Sprachbunds: A view from the Balkans. *The Cambridge Handbook of Areal Linguistics*, **55**, 87.

Friedman, V. A., & Joseph, B. D. (2025). *The Balkan Languages*. Cambridge University Press.

Georgiou, G. P. (2021). Toward a new model for speech perception: The Universal Perceptual Model (UPM) of Second Language. *Cognitive Processing*, **22**(2), 277–289.

Georgiou, G. P. (2023). Speakers of different L1 dialects with acoustically proximal vowel systems present with similar nonnative speech perception abilities: Data from Greek listeners of Dutch. *Speech Communication*, **150**, 32–40.

Georgiou, G. P. (2024). Classification of English vowels in terms of Cypriot Greek categories: the role of acoustic similarity between L1 and L2 sounds. *Canadian Journal of Linguistics*, **69**(1), 46–62.

Georgiou, G. P. (2025a). Leveraging a deep learning neural network for classifying native and nonnative speakers based on acoustic features. In I. Spantidakis, N. Dinas, V. Chadjinikita, & E. Griva (Eds.), *Γλώσσα, Εκπαίδευση και Τεχνητή Νοημοσύνη [Language, Education, and Artificial Intelligence]* (pp. 205–214). University of Crete.

Georgiou, G. P. (2025b). Enhancing developmental language disorder identification with artificial intelligence: Development of an explainable screening app using real and synthetic data. *Journal of Autism and Developmental Disorders*. doi: 10.1007/s10803-025-07176-1

Georgiou, G. P., & Dimitriou, D. (2023). Perception of Dutch vowels by Cypriot Greek listeners: To what extent can listeners' patterns be predicted by acoustic and perceptual similarity? *Attention, Perception, and Psychophysics*, **85**, 2459–2474.

Georgiou, G. P. & Giannakou, A. (2024). Acoustic characteristics of Greek vowels produced by adult heritage speakers of Albanian. *Acoustics*, **6**, 257–271.

Georgiou, G. P., Giannakou, A., & Alexander, K. (2025). Perception of second language phonetic contrasts by monolinguals and bidialectals: A comparison of competencies. *Quarterly Journal of Experimental Psychology*, **78**(6), 1148–1162.

Georgiou, G. P., & Paphiti, M. (2026). Machine learning classification of autistic adults using vowel acoustic features. *Research in Autism*, **137**, 202995.

Georgiou, G. P., & Savva, E. (2026). Exploring acoustic overlap in second language vowel productions. *International Journal of Applied Linguistics,* **36**(2), 1293–1304.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, **97**(5), 3099–3111.

Idrisoglu, A., Moraes, A. L. D., Cheddad, A., Anderberg, P., Jakobsson, A., & Berglund, J. S. (2025). Vowel segmentation impact on machine learning classification for chronic obstructive pulmonary disease. *Scientific Reports*, **15**(1), 9930.

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press

Jibson, J. (2022). Formant detail needed for identifying, rating, and discriminating vowels in Wisconsin English. *The Journal of the Acoustical Society of America*, **151**(6), 4004–4015.

Joseph, B. D. (2003). *When Languages Collide: Perspectives on Language Conflict, Language Competition, and Language Coexistence*. Ohio State University Press.

Kaskampa, A., & Georgiou, G. P. (2025). An examination of Greek vowel production by adult Albanian speakers of Greek as a second language. In C. Savvidou, G. P. Georgiou, & A. Giannakou (Eds.), *Advances in second and foreign language acquisition: From Greek classrooms to global contexts* (pp. 33–48). Palgrave Macmillan.

Lengeris, A. (2009). Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese. *Phonetica*, **66**(3), 169–187.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, **48**(4), 839–862.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, **49**(2B), 606–608.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, **30**, 4765–4774.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge University Press.

McCarthy, R. A., Zhang, Y., Verburg, S. A., Jenkins, W. F., & Gerstoft, P. (2025). Machine learning in acoustics: A review and open-source repository. *npj Acoustics*, **1**(1), 18.

Millet, J., Chitoran, I., & Dunbar, E. (2025). *Gradient Phonological Overlap, Not Categorical Resolution, Predicts Speech Sound Discrimination*. Available at SSRN: https://ssrn.com/abstract=5499247 or https://doi.org/10.2139/ssrn.5499247

Mora, J. C., & Fullana, N. (2007). Production and perception of English/i:/-/I/and/æ/-/^/in a formal setting: Investigating the effects of experience and starting age. In Proceedings of the 16th international congress of phonetic sciences (Vol. 3, pp. 1613–1616). Saarbrücken: Universität des Saarlandes.

Munro, M. J. (2022). Variability in L2 Vowel production: Different elicitation methods affect individual speakers differently. *Frontiers in Psychology*, **13**, 916736.

Nittrouer, S. (2007). Dynamic spectral structure specifies vowels for children and adults. *The Journal of the Acoustical Society of America*, **122**(4), 2328–2339.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, **24**(2), 175–184.

R Core Team (2026). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Recasens, D., & Espinosa, A. (2009). Dispersion and variability in Catalan five and six peripheral vowel systems. *Speech Communication*, **51**(3), 240–258.

Schwartz, J. L., Boë, L. J., Vallée, N., & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, **25**(3), 255–286.

Song, J. Y., & Eckman, F. (2025). The relationship between L2 learners' production and perception of English vowels: The role of native-speaker acoustic patterns in production. *Second Language Research*, **41**(2), 309–338.

Themistocleous, C. (2019). Dialect classification from a single sonorant sound using deep neural networks. *Frontiers in Communication*, **4**, 64.

Topintzi, N. (2023). *Phonetics and phonology: An introduction, with special focus on English and Greek*. Kallipos, Open Academic Editions.

Van Leussen, J. W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, **6**, 1000.

Williams, D., & Escudero, P. (2014). Influences of listeners' native and other dialects on cross-language vowel perception. *Frontiers in Psychology*, **5**, 1065.

Williams, D., Escudero, P., & Gafos, A. (2018). Spectral change and duration as cues in Australian English listeners' front vowel categorization. *The Journal of the Acoustical Society of America*, **144**(3), EL215–EL221.

Cite this article: Georgiou, G. P. & Giannakou, A. (2026). Modeling cross-language vowel similarity: integrating phonetic analysis and explainable classification. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716426101076

https://doi.org/10.1017/S0142716426101076 Published online by Cambridge University Press