Learning to produce and transcribe cardinal vowels: speech and language therapy students' perception of task difficulty

Nicole Whitworth

Published online: 16 Apr 2024.

Submit your article to this journal ⤤

Article views: 380

View related articles ⤤

View Crossmark data ⤤

Taylor & Francis
Taylor & Francis Group

Learning to produce and transcribe cardinal vowels: speech and language therapy students' perception of task difficulty

Nicole Whitworth (ID)

Carnegie School of Education, Leeds Beckett University, Leeds, UK

ABSTRACT

Learning vowel transcription skills is crucial to function as a Speech and Language Therapist (SLT). However, vowel transcription is commonly regarded as particularly difficult and therefore often avoided. Despite the importance of accurate transcriptions, little is known about all the factors that influence the process of learning vowel transcription, which usually includes the learning of the Cardinal Vowel (CV) system. There are only a few studies that investigate how CVs are learnt and what factors lead to successful learning. The current study reports students' perceived difficulty of producing and transcribing CVs as a first step to identify how perceived difficulty affects phonetic learning. Perceived difficulty ratings for the production and transcription of 12 CVs collected from 155 students studying towards a qualification as an SLT were analysed. The results show that the classificatory features correlate with the perceived task difficulty of production and transcription. Implications for teaching are outlined.

Introduction

Phonetic transcription skills are an essential component of becoming and being a competent speech and language therapist (Howard & Heselwood, 2002; Knight et al., 2018; Shaw & Yanushevskaya, 2022; Titterington & Bates, 2021). Phonetic transcription is a complex task that requires the transcriber to link sound, symbol and articulatory configurations whilst adjusting for speaker variation as well as prosodic and phonetic context. Transcription requires the user to produce as well as recognise the sound auditorily and assign the appropriate symbol. The process of acquiring these skills begins during SLT students' training at university in practical phonetics classes and some level of continued training is required throughout clinicians' working life (Knight et al., 2018; Shaw & Yanushevskaya, 2022). One aspect of practical phonetics training is learning to produce and transcribe vowel sounds (Knight et al., 2014). Whilst learning phonetic transcription generally is considered a challenging task, transcribing vowels is usually considered even more of a challenge (Shaw & Yanushevskaya, 2022; Titterington & Bates, 2018). A first step in mastering vowel transcription is for students to learn to produce and transcribe Cardinal Vowels (CVs). This paper is concerned with assessing students' perception of the difficulty of producing and transcribing CVs.

The research reported in this paper increases our understanding of SLT students' perceived difficulty of producing and transcribing CVs. On a practical level, this will assist in the design of suitable methods to teach CVs in a way that facilitates SLTs' use of vowel transcription of clinical speech in an environment where there are increasing time constraints on the number of hours that can be devoted to the learning of phonetic transcription skills (Knight et al., 2018; Titterington & Bates, 2018, 2021). More widely, as Mackenzie-Beck (2003) speculates, a better understanding of phonetic learning may also shed light on how sounds are processed and learned by the human mind. The results therefore may contribute to a better understanding in other areas of phonetics, for example, acquisition of L2 sounds and speech perception and production generally.

Importance of transcription skills for clinical practice

The importance of phonetic transcription is undisputed in the literature, even though it is generally acknowledged that in clinical practice there are often time-constraints preventing their universal use (Howard & Heselwood, 2002; Mackenzie-Beck, 2003). Phonetic transcription skills are essential not just for analysis, diagnosis and the identification and management of a suitable treatment plan, but the learning of these skills fosters in the future speech and language therapist a better understanding of how speech and language work (Howard & Heselwood, 2002; Shaw & Yanushevskaya, 2022; White et al., 2022). Phonetic transcription is arguably more important for clinicians working with some clinical populations, e.g. individuals with cochlear implants (Teoh & Chin, 2009), children with SSD (Nelson et al., 2019), etc. However, since SLTs typically qualify to work across the spectrum, it is a compulsory part of the curriculum of SLT courses (Knight et al., 2018).

The importance of accurate vowel transcription is highlighted in the Child Speech Disorder Research Network (2017) *Good Practice Guidelines* (p. 5) alluding to a tendency to focus on the transcription of consonants and neglect vowels whose transcription is considered more difficult. The relevance of providing accurate, detailed transcriptions of vowels is echoed by other scholars such as Ball et al. (2010), Howard and Heselwood (2002), Howard and Heselwood (2013), Pollock and Berni (2001), Teoh and Chin (2009), not least because the effect of mismatches in vowel articulation on intelligibility is significant (Teoh & Chin, 2009) and can be indicative of more complex Speech Sound Disorder (SSD) which makes up a large proportion of SLT workload (Child Speech Disorder Research Network, 2017).

Learning transcription skills

Overviews of the challenges of teaching and learning phonetic transcription for clinical practice have previously been provided by various authors, e.g. Ball et al. (2009), Howard and Heselwood (2002), Stemberger and Bernhardt (2019). Despite its importance, comparatively little is known about how transcription is learnt and what factors affect successful learning of phonetic transcription. Studies have mainly looked at which learner characteristics facilitate the acquisition of transcription skills: Success in learning to transcribe phonetically has been shown to be affected by the learner's phonological awareness (Moran & Fitch, 2001; Robinson et al., 2011), short-term memory (Knight & Maguire, 2011), musical aptitude (Mackenzie-Beck, 2003) and the ability to identify one's own tongue

position (Mackenzie-Beck, 2003). Mackenzie-Beck (2003) also suggested that teaching can overcome some of the initial lack-of-aptitude disadvantages in the long run. Knight (2010) showed that accuracy in phonetic transcription assessment is improved by different voices and increased number of repetitions.

Transcribing vowels

In comparison with consonants, vowels are universally considered more difficult to classify, analyse and transcribe (Howard & Heselwood, 2013; Teoh & Chin, 2009). Practicing SLTs report that they find vowel transcription confusing, calling for more opportunities to practice it (Knight et al., 2018; Nelson et al., 2019; Titterington & Bates, 2018). There are several commonly cited reasons for this: the lack of fixed points of reference in terms of their articulation, greater articulatory variability, reduced distinctness between perceptual categories and the use of different notation systems and transcription methods (Nelson et al., 2019; Pollock & Berni, 2001).

In the transcription of clinical speech, there are two main ways of transcribing vowels: The first is to use the accent-specific accepted realisation of the target phoneme and suitable diacritics (e.g. lowered, raised, advanced, etc.) to show how the vowel that is to be transcribed diverges from it. To make such a transcription interpretable by others, this requires a note showing the expected values for the relevant accent using CVs (Howard & Heselwood, 2013). The second method is to use CVs as reference vowel qualities, so that what is recorded is the closest CV plus any diacritics to show how the vowel that is being transcribed differs from it in terms of tongue position and lip rounding. Howard and Heselwood (2013, pp. 87–90) discuss the advantages and disadvantages of both approaches. Important for the discussion here is that to use either method, it is necessary to assess the position of the highest point of the tongue within the vowel space in the oral cavity and make relatively fine-grained distinctions between the target articulation and the actual articulation.

Learning the principles of the CV system is thus an essential preparation for vowel transcription. Despite well-documented issues with the CV system and its use, in practice, it offers a suitable technique for transcribing vowels (Abercrombie, 1967; Howard & Heselwood, 2013). Not all students learn the whole cardinal vowel system, as a sub-set of these is sufficient for describing and transcribing speech (Abercrombie, 1967, p. 160). Students on SLT courses typically only learn the full set of eight Primary Cardinal Vowels (PCVs), namely CV1 [i], CV2 [e], CV3 [ɛ], CV4 [a], CV5 [ɑ], CV6 [ɔ], CV7 [o] and CV8 [u], and four of the Secondary Cardinal Vowels (SCVs), namely CV9 [y], CV10 [ø], CV11 [œ] and CV16 [ɯ] (Ashby, 2002, 2003; Knight, 2010; Wikström & Setter, 2011).

Learning the cardinal vowels

There are several publications that provide recommendations on how best to teach IPA sounds including vowels (Howard & Heselwood, 2002; Knight et al., 2014, 2021), but these are usually based on teacher experience rather than empirical data. As for all ear-training classes, teaching is ideally carried out in small groups which permits students to practise production and transcription of the vowels. Generally, only a comparatively small amount of time can be devoted to the acquisition of the CVs and students are expected to practice in

their own time using self-study materials. As part of the learning, students are typically given live and recorded examples of each of the CVs and are asked to produce these. Feedback on these productions and suitable articulatory instructions are used to help students achieve the desired auditory effect. Howard and Heselwood (2002) give the following example:

> For cardinal vowels, first-language vowels can be used as starting points with the instruction to exaggerate them and make them more extreme. Where the combination of tongue-position and lip-shape is unfamiliar, instructing the student to adopt the lip-shape for one vowel but to 'think' another vowel with the same tongue position is often successful. Cardinal 16 [ɯ] will usually be reasonably satisfactorily produced with the instruction to put the lips in the shape for [i] but to 'think [u]'.

Whilst there is a wealth of advice based on teacher experience, as evidenced in the exercises provided in, for example, Ladefoged (2001, pp. 204–206) and Catford (2001, pp. 120–152), there are only three empirical studies that have looked at how individuals learn the cardinal vowel system and these look at different aspects from different angles.

Ashby (2002, 2003) investigated students' responses in phonetic transcription assessments of selected IPA sounds including CVs at two points in time. Her participants consisted of 125 students studying phonetics as part of the first year of a joint honours UG degree in linguistics. The data were collected from five successive cohorts of students. Overall, 43% were non-native speakers of English, and 57% were native speakers of English. Students were given tests 12 weeks and 24 weeks after commencing their general phonetic training encompassing 1 h of ear-training per week. CVs were presented in isolation and as part of nonsense words. The study found for the set of PCVs, that some CVs appeared to be easier to learn based on statistically significantly different rates for correct transcription (Ashby, 2002, pp. 231–234). The study found that errors occurred most often with regard to the vertical (open-to-close) plane. In the initial test, incorrect responses were received most frequently (79%) for height-adjacent vowels [i]-[e], [e]-[ɛ], [ɛ]-[a], [u]-[o] and [o]-[ɔ]. On inclusion of SCVs in a later test only 15% of incorrect responses for the SCVs were attributable to tongue height.

Two unpublished conference presentations report exploratory data on the perceived difficulty of IPA sounds: Whitworth (2008) investigated the perceived difficulty of IPA sounds including CVs for 86 students on an SLT course. She found that PCVs were generally judged to be easier to produce and transcribe than SCVs. PCVs were judged to be difficult where their tongue height did not correspond to that usually found in English. SCVs were judged to be more difficult where they have a lip posture/tongue position combination not found in English. Overall, students rated more sounds as difficult to transcribe than difficult to produce. Whitworth (2011) looked more specifically at the relationship between the number of CVs SLT students rated as difficult and their exam performance for CV production and perception, including the responses from an additional 24 students. She reports a moderate positive correlation between the number of sounds perceived to be difficult and a higher exam mark. No data was given as to the statistical significance of this correlation. Considering the variability in marks for students who have rated the same number of vowels as difficult, it is likely that the correlation is not significant but random.

Wikström and Setter (2011) assessed self-reported confidence (on a scale of 1–4 per CV) and performance (test scores at two points in time) for CV perception and production of six students enrolled on a clinical phonetics and phonology module. All speakers were monolingual speakers who were phonetically untrained prior to starting the module. They reported that in the transcription task, students were able to identify PCVs reliably with the surprising exception of [ɑ]. SCVs were found to be a source of errors. The CV pairs [o] and [ɔ] as well as [u] and [y] were shown to be confused by their participants. Overall, their study identified height as most problematic followed by roundedness for transcription tasks. For production, this study assessed students' productions acoustically by comparing it to those of the two teachers. They found that differences in vowel height also posed the greatest problems for production, such as students producing a more open vowel for [e] and [o]. They note further that front rounded and back unrounded vowels were challenging for all students to produce. The authors attributed this to the students' L1 English backgrounds. Their analysis of the self-reported confidence levels indicated that there is a link between how confident students feel about producing and transcribing vowels and how they actually performed in the assessments, but they also acknowledge that for some sounds students were overly and for others insufficiently confident.

All studies report some agreement that vowel height differences are challenging both in transcription and production, with open-mid and close-mid being the most difficult ones. The backness and rounding dimensions are deemed difficult mostly for SCVs. However, the definition of difficulty is not the same in the three studies: Ashby (2002, 2003) established CV difficulty based on an error analysis of student transcriptions but did not investigate how the students' perceived difficulty of the tasks of transcribing and producing these sounds. Whitworth (2011) and Wikström and Setter (2011) looked at both how students felt about CV production and transcription and their actual performance. Whilst both report a moderate correlation between sounds perceived to be difficult and actual performance, Whitworth (2011), did not provide inferential statistics and the sample size in Wikström and Setter (2011) was very small and therefore not representative.

Relevance of perceived task difficulty to teaching and learning

Teaching and learning literature recognises the importance of understanding perceived task difficulty in the learning process (Chen et al., 2022; Schneider et al., 2022; Stephanou et al., 2011; Street et al., 2022). Perceived task difficulty is defined as a person's subjective judgement of the amount of effort required to complete the task and the likelihood of completing it successfully (Andrabi et al., 2022). It is influenced by several factors related to the person's prior experiences, personality traits and motivation (Andrabi et al., 2022). Perceived task difficulty is known to generate positive or negative emotions (e.g. boredom, confidence and/or enthusiasm) which determine a student's level of engagement (Chen et al., 2022; Schneider et al., 2022; Stephanou et al., 2011). It is correlated with self-regulation in the achievement of academic goals and therefore is a factor in academic achievement (Stephanou et al., 2011). Students who perceive a task to be more difficult typically perform worse in exams (Chen et al., 2022; Stephanou et al., 2011). However, teacher support (both cognitive and emotional) can mitigate negative effects where a task is perceived to be difficult (Chen et al., 2022; Street et al., 2022). In addition, Schneider et al.

(2022) found that increasing learner autonomy by providing a choice of learning task can reduce the effect of high levels of perceived task difficulty on learning achievement.

As has already been outlined above, one of the recurring themes of the clinical phonetics teaching literature is that phonetics and particularly vowel transcription and production are perceived to be difficult. Howard and Heselwood (2002) specifically note that 'all sounds are not, indeed, equal when it comes to ease and accuracy of production, nor of transcription' (Howard & Heselwood, 2002, p. 379). They comment on the remarkable consensus among students and which sounds are perceived to be difficult to transcribe or produce. Given the effect of perceived task difficulty on learning outcomes as outlined above, and the importance of successful learning of phonetics by future speech and language therapists, it is therefore important that teachers of phonetics have a good understanding of which sounds are likely to be perceived to be difficult to produce and transcribe. Whilst all phonetics teachers undoubtedly have professional insight and intuition, there is no study that investigates this based on a large set of data.

Aims of this study

Building on the unpublished studies by Whitworth (2008, 2011), this study establishes which sounds students judge to be difficult to produced and transcribe using secondary data collected from a large number of SLT students. In addition to what has been reported before, this study identifies and compares the features of the sounds that are perceived to be challenging to produce and transcribe and examines the relationship between perceived difficulty of perception and production. It constitutes an important first step towards understanding the learning and teaching of CV sounds in the phonetics classroom. Results will be valuable for teachers who wish to base the design and delivery of course materials on evidence-based insights rather than intuition and will add to our understanding of the learning and teaching of the sounds of the IPA. This paper reports the results for Cardinal Vowels. Results for the IPA consonants covered in the syllabus are covered in a separate publication.

Research questions

RQ1: What, if any, classificatory features of CVs affect SLT students' ratings of CVs as 'difficult to produce' or 'difficult to transcribe'?

RQ2: Is there a hierarchy of difficulty for CVs considering perceived difficulty to produce and perceived difficulty to transcribe ratings?

Materials and methods

Materials and data collection

The study used secondary data that had been collected anonymously at the end of the academic year to inform the module revision session. Ethical approval for this study was

granted through the Leeds Beckett University ethical approval procedure (application number 117103).

At the time of data collection, the participants had been taught transcription and production of a set of sounds taken from the standard International Phonetic Alphabet that are relevant in the transcription of clinical speech. Specifically, students had been taught 2 hour/s a week of general phonetic theory and 1 hour/s per week of practical phonetics in small groups (<15 students) for 24 weeks, resulting in a total of 48 hour/s of theoretical and 24 hour/s of practical phonetics. The students were from four cohorts enrolled on the same module in four consecutive years prior to 2016. There were no repeating students. All students were taught by the same teacher, a trained phonetician with over 10 years' experience of teaching phonetics to speech and language therapy students in UK Higher Education, using the same programme. In addition to face-to-face classes, students' development had been supported by weekly online materials including videos, sound files and tasks. For vowels, theoretical concepts and practical transcription had been allocated a total of 6 hour/s of which two were practical.

The data had been collected using two paper-based anonymous questionnaires from 155 students who had completed the module as outlined above. Each questionnaire was laid out in the same way. One side listed all IPA consonant sounds on the syllabus in tabular form as set out on the standard IPA chart. On the reverse, all CVs on the syllabus were listed in tabular form including number and symbol. For reference purposes, a labelled vowel quadrilateral was also provided. Students were first given the questionnaire asking them to circle all the sounds that they considered 'difficult to produce'. They were then given the second questionnaire, and were instructed to circle all sounds that they considered 'difficult to transcribe'. If students were unsure about a symbol, they could ask for clarification. The collated questionnaire sets (comprising one 'difficult to produce' and one 'difficult to transcribe' questionnaire per student) were then collected from the students.

Data analysis

To prepare the data for analysis, the student responses were collated in an Excel spreadsheet, recording a 1 where a sound had been circled (indicating it was considered difficult) and a 0 where a sound was not circled. In addition, for each CV, counts of the number of times it was rated as difficult in the two modes were made, resulting in two measures: PDP representing the number of times a CV was rated as 'difficult to produce', and PDT representing the number of times a CV was rated as 'difficult to transcribe'. Each CV was also classified in articulatory terms as shown in Table 1.

Of 155 questionnaire sets, 16 were excluded due to responses being unclear, e.g. a circle may or may not have been crossed out, etc. The remaining 139 questionnaire sets were included in the analysis, comprising a total of 3336 data points. Table 1 shows the classification of each CV and the counts and the percent of PDP (Perceived 'Difficult to Produce') and PDT (Perceived 'Difficult to Transcribe').

All statistical analyses were performed in R version 4.3.2 (The R Foundation, 2023) using RStudio version 2023.06.0 (Posit Software, 2023). As a first step, chi-square was used to assess the statistical significance of the differences between PDT and PDP ratings for each CV. Then, multiple logistic regression modelling (Diez et al., 2019, pp. 371–378) was used to fit the data based on the binary dependent variable difficult/not_difficult. The independent

Table 1. PDP and PDT counts and percent by CV and CV articulatory classification used in the logistic regression modelling.

CV		Front	Open	Lips	CVSet	PDP	%PDP	PDT	%PDT
1	[i]	FRONT	CLOSE	SPREAD	PRIMARY	0	0%	4	3%
2	[e]	FRONT	CLOSE-MID	SPREAD	PRIMARY	39	28%	50	36%
3	[ɛ]	FRONT	OPEN-MID	NEUTRAL	PRIMARY	19	14%	43	31%
4	[a]	FRONT	OPEN	NEUTRAL	PRIMARY	4	3%	15	11%
5	[ɑ]	BACK	OPEN	NEUTRAL	PRIMARY	14	10%	20	14%
6	[ɔ]	BACK	OPEN-MID	ROUNDED	PRIMARY	32	23%	51	37%
7	[o]	BACK	CLOSE-MID	ROUNDED	PRIMARY	37	27%	58	42%
8	[u]	BACK	CLOSE	ROUNDED	PRIMARY	12	9%	18	13%
9	[y]	FRONT	CLOSE	ROUNDED	SECONDARY	34	24%	24	17%
10	[ø]	FRONT	CLOSE-MID	ROUNDED	SECONDARY	70	50%	73	53%
11	[œ]	FRONT	OPEN-MID	ROUNDED	SECONDARY	65	47%	66	47%
16	[ɯ]	BACK	CLOSE	SPREAD	SECONDARY	74	53%	26	19%

variables examined in the full model were frontness (front/back), openness (close/close-mid/open-mid/open), lip posture (spread/neutral/rounded) and CVSet (primary/secondary). After the full model was fitted, backward stepwise logistic regression was used to identify a simplified model. To exclude multicollinearity, the variance inflation factor was calculated for all models. Statistical significance of the models was determined with log-likelihood ratios and the Akaike Information Criterion (AIC) by comparing the Null model to the simplified and full models (Diez et al., 2019, p. 374). Finally, to group the CVs into clusters of perceived difficulty using PDP and PDT, a hierarchical clustering analysis was carried out to develop a hierarchy of difficulty.

Results

Perceived 'difficult to produce' (PDP)

In total, CVs were rated as 'difficult to produce' 400 times (24%). Figure 1 shows the distribution of PDP ratings for individual CVs in decreasing order.

A chi-square (χ^2) test of independence was performed to test the relationship between PDP and CV quality. The relationship between these variables was highly significant, χ^2 ($df = 11$, $N = 1764$) = 290.9, $p < 0.001$.

Analysis of PDP by articulatory features

The boxplots in Figure 2 show the descriptive statistics for PDP.

Logistic regression was used to analyse the relationship between lip posture, frontness, openness and the number of CVs rated as 'difficult to produce' in comparison to the reference level CV1. The results are presented in Table 2. Figure 3 illustrates the Odds Ratios.

Holding all other predictor variables constant, CVs classified as back were 2.59 times (95% CI [0.64, 1.26], $p < 0.001$) more likely to be rated as 'difficult to produce'. CVs classified as close-mid were 4.9 times (95% CI [1.22, 1.98], $p < 0.001$) and those classified as open-mid 3.63 times (95% CI [0.88, 1.71], $p < 0.001$) more likely to be rated as 'difficult to produce'. The likelihood of a CV to be PDP for open CVs was not significantly different (95% CI [−0.91, 0.69], $p = 0.79$) from the reference value close. The likelihood of a CV to be PDP for CVs with a neutral lip posture was not significantly different (95% CI [−0.98, 0.38], $p = 0.41$) from the
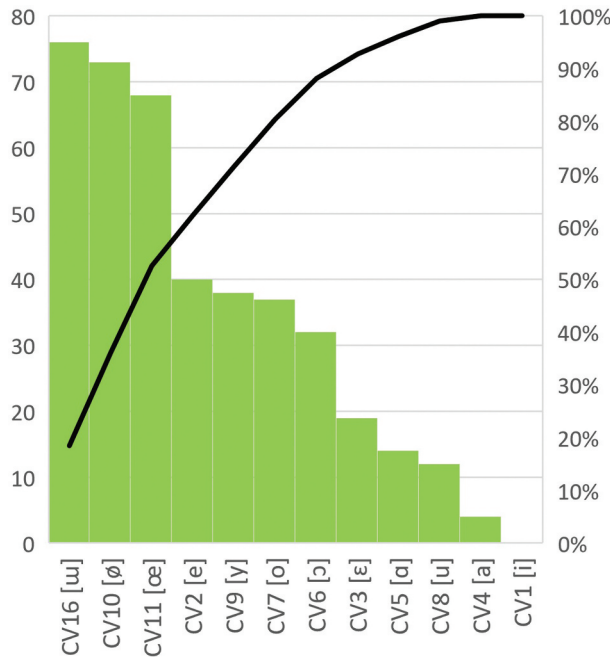
Figure 1. Mean PDP for production for each CV in decreasing order. The pareto line shows cumulative PD.

reference value spread. CVs classified as rounded were 0.54 (95% CI [−0.95, −0.28], $p < 0.001$) times less likely to be rated as 'difficult to produce' than the reference value spread (SPR). CVs belonging to the secondary set of CVs were 7.77 times (95% CI [1.72, 2.38], $p < 0.001$) more likely to be rated as 'difficult to produce' than those belonging to the primary CVs.

A simplified model was therefore constructed without the lip posture parameter which was found to be only of limited significance. Table 3 shows the full model above and a simplified model that excludes lip posture are a similar fit with lower levels of collinearity for the simplified model. Both models are significantly different from the Null Model. The most pertinent features that explain students' PDP judgements are CVSet, openness, and frontness.

Perceived 'difficult to transcribe' (PDT)

In total, CVs were rated as 'difficult to transcribe' 448 times (24%). Figure 4 shows the distribution of PDT ratings for individual CVs in decreasing order.

A chi-square (χ^2) test of independence was performed to test the relationship between PDT and CV quality. The relationship between these variables was highly significant, χ^2 ($df = 11$, $N = 1752$) $= 207.4$, $p < 0.001$.

Analysis of PDT by articulatory features

The boxplots in Figure 5 show the descriptive statistics for PDT.

Logistic regression was used to analyse the relationship between lip posture, frontness, openness and the number of CVs rated as 'difficult to transcribe' (PDT) in
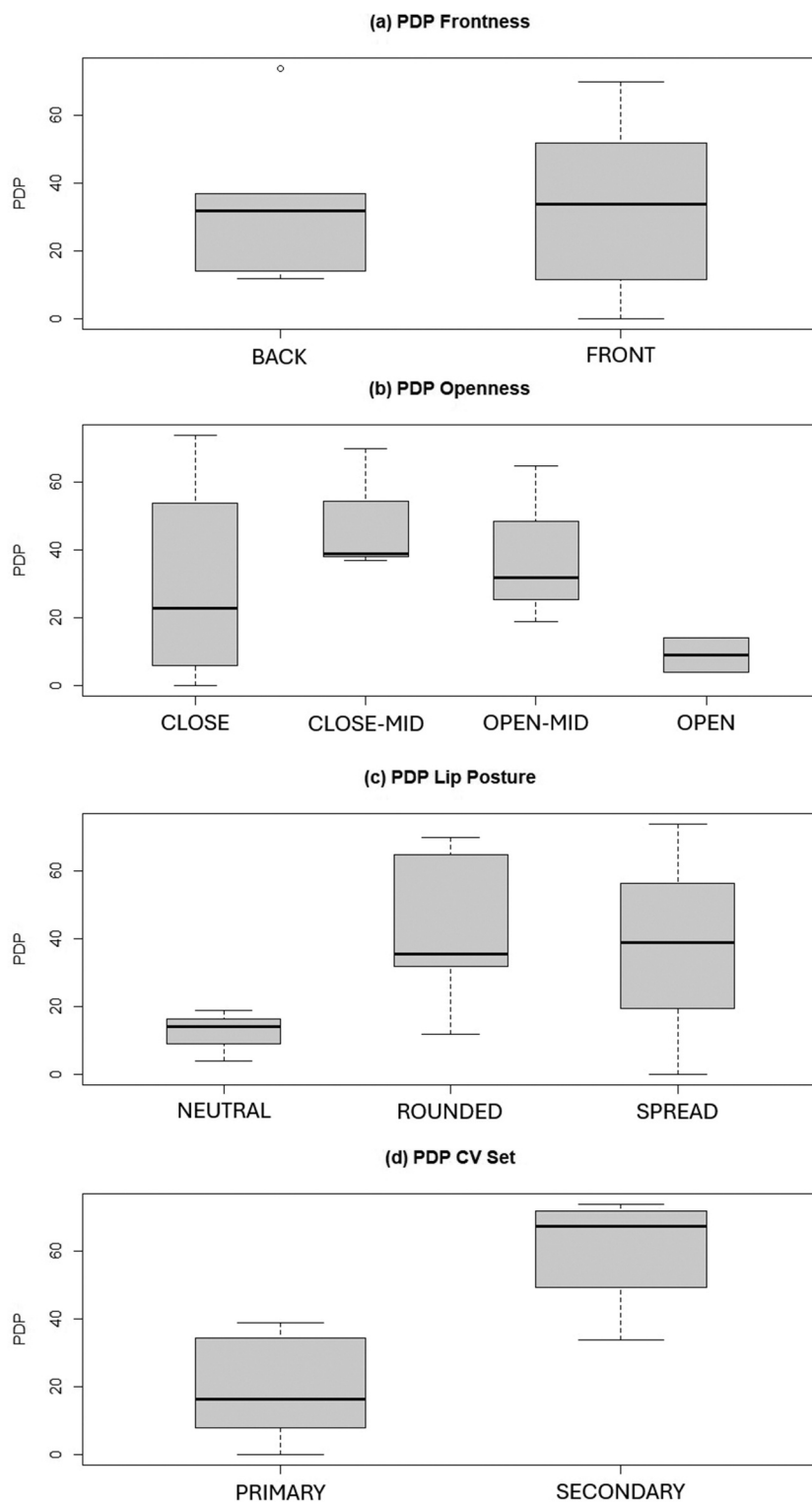
Figure 2. Boxplots of PDP by articulatory classification: (a) frontness, (b) openness, (c) lip posture and (d) CVSet.

Table 2. Output of logistic regression full model for PDP. The intercept is the front close spread CV1 [i].

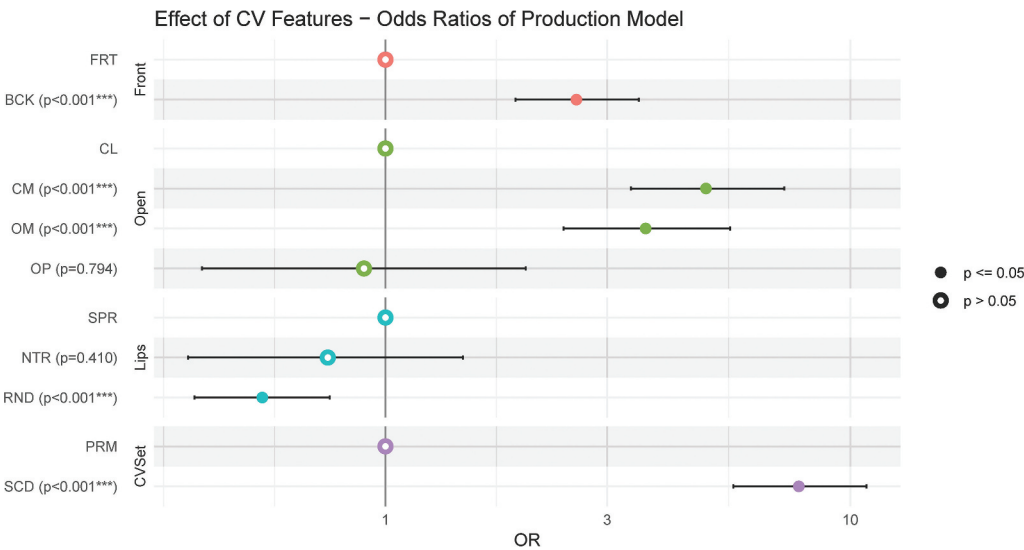|	Log Odds (β)	Odds Ratio (e^β)	SE	z-value	p-value
Intercept	−2.85	0.06	0.29	−12.44	<0.001
Front-BACK	0.95	2.59	0.16	6.08	<0.001
Open-CLOSE-MID	1.59	4.90	0.19	8.20	<0.001
Open-OPEN-MID	1.29	3.63	0.21	6.13	<0.001
Open-OPEN	−0.11	0.90	0.41	−0.26	0.79
Lips-NEUTRAL	−0.29	0.74	0.35	−0.83	0.41
Lips-ROUNDED	−0.61	0.54	0.17	−3.57	<0.001
CVSet-SECONDARY	2.05	7.77	0.17	12.18	<0.001

Figure 3. Odds ratio plot illustrating the effect of CV features on PDP in the full model. FRT = front, BCK = back, CL = close, CM = close-mid, OM = open-mid, OP = open, SPR = spread, NTR = neutral, RND = rounded, PRM = primary CV, SCD = secondary CV.

comparison to the reference level CV1. The results are presented in Table 4. Figure 6 illustrates the Odds Ratios.

Holding all other predictor variables constant, CVs classified as back were 1.42 times (95% CI [0.05, 0.66], $p < 0.001$) more likely to be rated as 'difficult to transcribe'. CVs classified as close-mid were 6.42 times (95% CI [1.51, 2.24], $p < 0.001$) and those classified as open-mid (OM) 4.85 times (95% CI [1.19, 1.98], $p < 0.001$) more likely to be rated as 'difficult to transcribe'. The likelihood of a CV to be PDT for open CVs was not significantly different (95% CI [−0.39, 0.90], $p = 0.43$) from the reference value close. The likelihood of a CV to be PDT for CVs with a neutral lip posture was not significantly different (95% CI −0.34, 0.80], $p = 0.42$) from the reference value spread. CVs classified as rounded were also not significantly more likely (95% CI [−0.20, 0.49], $p < 0.42$) to be rated difficult as compared to the reference value. CVs belonging to the secondary set of CVs were 2.2 times (95% CI [0.46, 1.12], $p < 0.001$) more likely to be rated as 'difficult to transcribe' than primary CVs.
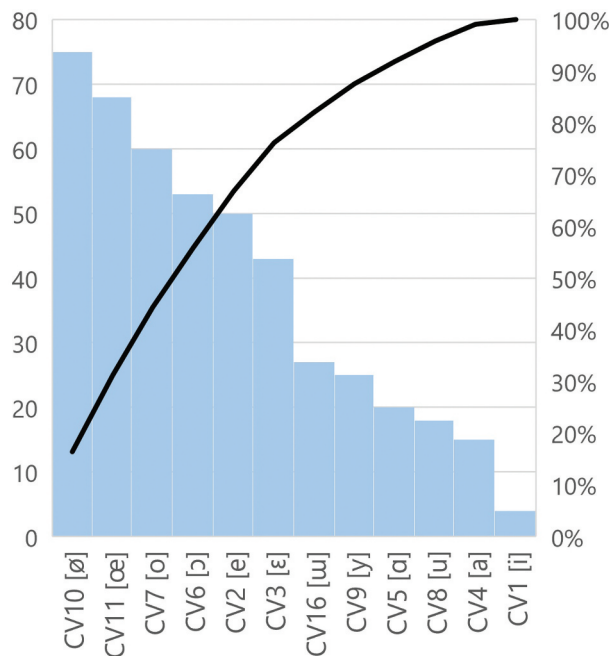
Figure 4. Mean PDT for production for each CV in decreasing order. The pareto line shows cumulative PDT.

Table 3. Comparison of null model, simplified model and full model for PDP.

	AIC	LogLik	dF	χ^2 p-value	VIF
Null model	1839.7	−918.83	N/A	<0.001	N/A
Full model	1565.5	−774.77	7	<0.001	Front 1.49
					Open 4.27
					Lips 3.77
					CVSet 1.78
Simplified model	1575.1	−781.54	5	<0.001	Front 1.35
					Open 1.38
					CVSet 1.51

Table 5 shows the full model for PDT and a simplified model that excludes lip posture which was found not to produce significant differences in PDT. The simplified model is a similar fit with lower levels of collinearity. Both models are significantly different from the Null Model. The most pertinent features that explain students PDT judgements are CVSet, openness, and frontness.

Hierarchical cluster analysis of PDP and PDT data

Figure 7 shows a dendrogram of the results of the hierarchical cluster analysis of the PDP and PDT data. CVs are initially subdivided into two groups: the primary corner vowels and all remaining CVs. The remaining CVs are then further subdivided into two main groups, resulting in three main groups overall.
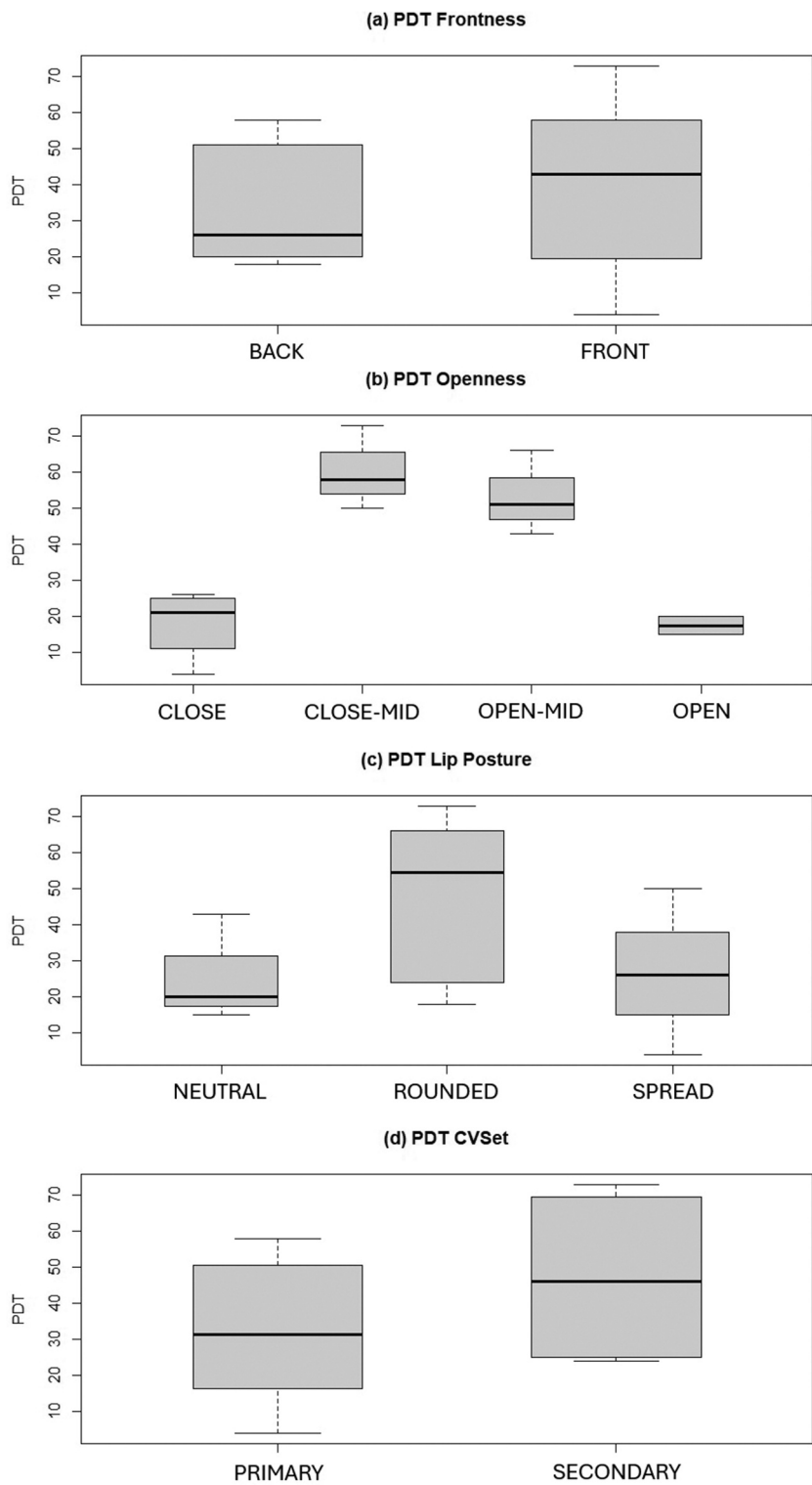
Figure 5. Boxplots of PDT by articulatory classification: **(a)** frontness, **(b)** openness, **(c)** lip posture, and **(d)** CVSet.

Table 4. Output of logistic regression full model for PDT. The intercept is the front close spread CV1 [i].

	Log Odds (β)	Odds Ratio (e^β)	SE	z-value	p-value
Intercept	−2.61	0.07	0.21	−12.54	<0.001
Front-BACK	0.35	1.42	0.15	2.30	0.02
Open-CLOSE-MID	1.86	6.42	0.18	10.10	<0.001
Open-OPEN-MID	1.58	4.85	0.20	7.81	<0.001
Open-OPEN	0.26	3.53	0.33	0.78	0.43
Lips-NEUTRAL	0.23	1.26	0.29	0.80	0.42
Lips-ROUNDED	0.14	1.15	0.18	0.80	0.42
CVSet-SECONDARY	0.79	2.20	0.17	4.68	<0.001
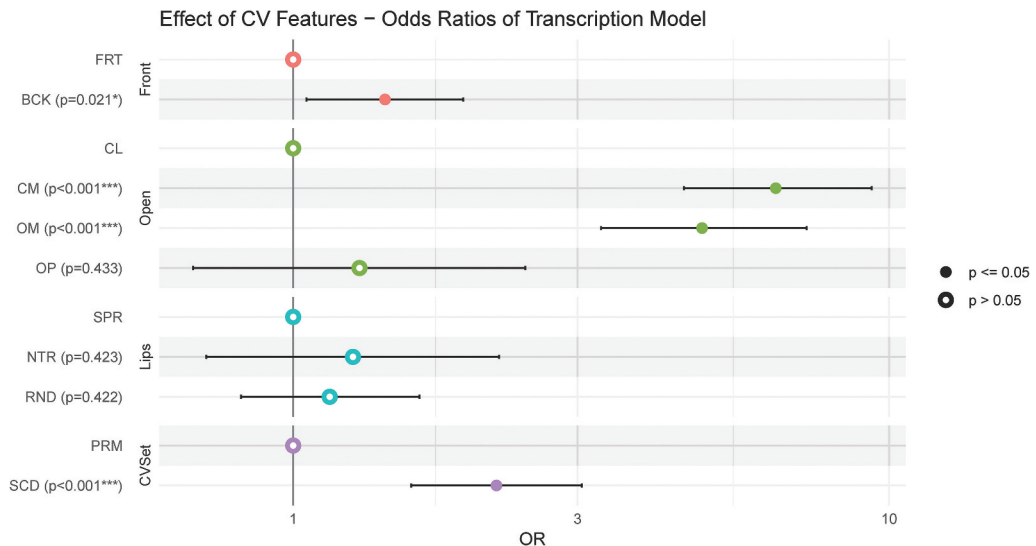
Figure 6. Odds ratio plot illustrating the effect of CV features on PDT in the full model. FRT = front, BCK = back, CL = close, CM = close-mid, OM = open-mid, OP = open, SPR = spread, NTR = neutral, RND = rounded, PRM = primary CV, SCD = secondary CV.

Table 5. Comparison of null model, simplified model, and full model for PDT.

	AIC	LL	df	χ² p-value	VIF
Null model		−970.52	N/A	<0.001	N/A
Full model	1747.4	−865.70	7	<0.001	Front 1.65
					Open 3.87
					Lips 4.33
					CVSet 1.90
Simplified model	1744.3	−866.14	5	<0.001	Front 1.22
					Open 1.25
					CVSet 1.34

Summary

The findings of this study show that student judgements of CVs as 'difficult to produce' and/or 'difficult to transcribe' correlate with classificatory labels. Specifically, for both task modes, secondary vowels are more difficult than primary ones, close-mid and open-mid CVs are more difficult than open and close CVs, and back vowels are more difficult than
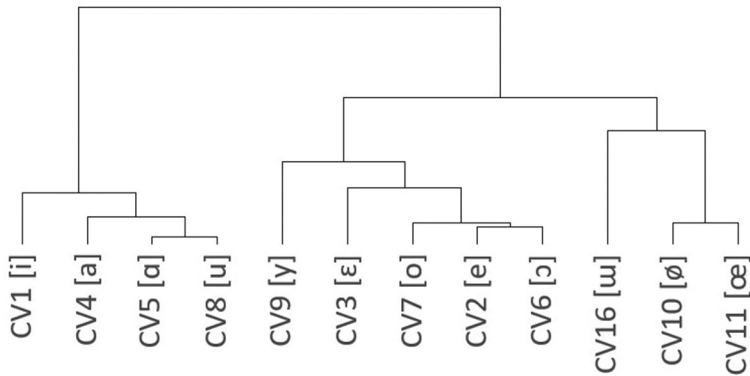
Figure 7. Cluster dendrogram.

front vowels. Lip posture does not appear to be a significant contributor to perceived difficulty. Rather, it appears that lip posture is more or less difficult depending on the accompanying horizontal tongue position captured in the design of the CV system by the primary vs. secondary CV dichotomy. The hierarchical cluster analysis grouped the CVs into two groups of perceived difficulty with primary CVs at the extreme corners of the vowel space as perceived to be least difficult.

The results correlate with findings from previous studies that identified that students experience most difficulties in transcription performance (Ashby, 2003) and in terms of their confidence and their performance (Wikström & Setter, 2011) for CVs in terms of vowel height differences as well as for secondary CVs. They add to the body of literature by providing a systematic large-scale examination of SLT students' perceived difficulty of CVs by classificatory feature.

Implications for teaching

Awareness of potential negative emotions caused by perceived task difficulty (PTD) may help the phonetics teacher to encourage ongoing engagement by giving them the opportunity to lower the perceived task difficulty and thereby help students achieve desired learning outcomes (Stephanou et al., 2011). Strategies to self-motivate and self-regulate can help students to develop meaning structures that help them deal with difficult learning situations. More specifically, being aware of students' PTD can help tutors to make decisions about the sequence and manner in which practical phonetics is taught.

Prior learning experience is a factor in shaping PTD judgements (Andrabi et al., 2022). To build confidence, teachers can provide students with positive experiences of production and transcription. Starting with sounds that have lower levels of PTD can increase students' confidence until they are willing to try more difficult tasks. Students can start by transcribing and producing the primary CVs located at the four corners of the vowel quadrilateral and then introducing CVs that have higher PTD.

At lower levels of perceived difficulty, emotional support through positive feedback can be useful to enhance student achievement (Chen et al., 2022). Where students have preconceived opinions of PTD or where they are actually experiencing difficulty, emotional

support alone is less effective (Chen et al., 2022). PTD ratings can be lowered by providing cognitive support (Chen et al., 2022) such as practical feedback that outlines specific and achievable actions in CV production and perception. For example, exercises that explore the dimensions of the vocal tract and the corresponding sounds that a specific articulatory posture produce alongside a live explanatory commentary of the relationship between changes in the vocal tract and sound output and vice versa. See Catford (2001) and Ladefoged (2001). It is good practice to encourage students to analyse their own production and transcriptions to understand whether a perceived difficulty is real and how to overcome it, providing a sense of autonomy that has been shown to lower PTD (Schneider et al., 2022). On a more general note, it is important to teach to recognise the fact that difficult tasks are worthwhile and can be accomplished to an acceptable level with practice and perseverance.

Limitations

The use of secondary data, whilst convenient and economical (Greenhoot & Dowsett, 2012), have of necessity introduced some limitations on the type of analyses that could be carried out. For example, one of the constraints introduced was a that the data was based on a binary decision task rather than eliciting a relative level of perceived difficulty. Binary judgements may result in either under- or overreporting of perceived difficulty not least due to random responding where the participant was either undecided or disengaged (Peng et al., 2023). On the other hand, a design requiring a more graded judgement of difficulty levels may have yielded lower-quality data due to the increased complexity of the choice task (Brown, 2016).

Another limitation was that since the data were collected anonymously to make sure that students felt confident to be honest about which CVs they perceived as difficult to produce or transcribe, it meant that correlations between perceived difficulty measures, learner characteristics, and learner achievement could not be explored. However, there are many ethical and efficiency benefits to using some of the vast amounts of data teachers collect to inform their teaching on a regular basis, at least for exploratory research purposes. For example, since the data were collected to serve the students' interest in highlighting sounds, they wanted to be reviewed in a revision session, it is likely that the ratings were more authentic than if they had been collected to inform a research project only.

Future research

Future research needs to look at how perceived difficulty of phonetic tasks affects learner success and how perceived task difficulty interacts with learner characteristics such as self-efficacy, anxiety, motivation, etc. Research in other areas of learning has shown that perceived task difficulty can compensate for task complexity (Yücel, 2022), improve learner performance (Street et al., 2022) and affected the impact of teacher support on student success (Chen et al., 2022). How far this applies to phonetic learning remains to be established.

Wikström and Setter (2011) found that students' confidence ratings, although linked to their actual performance, are not necessarily a good indicator of students' actual assessment performance. Students were both overly confident for some sounds and had low confidence for sounds that they performed well on in assessments. It would be of interest to understand

in more depth how and why students make specific difficulty ratings and what they represent in the perception of the students.

The study results mirror orders of vowel development in child speech (Kent & Miolo, 1995) and the occurrence of vowels in the world's languages (Ladefoged & Maddieson, 1996). At the same time, there seem to be some aspects that may be attributed to L1 influences. Evidence that L1 affects CV perception in trained phoneticians has been found in a study by Dioubina and Pfitzinger (2002). More research of the role of L1 background of the teachers and students in perceived difficulty ratings and the success of learning CVs is needed to tease out universal developmental from language-specific factors. Insights here will be useful in both second language teaching and the remedial aspects of speech and language therapy.

Conclusion

The results of the study confirm previous findings that students consider vowel height and front rounded and back unrounded vowels to be the most challenging for both production and transcription tasks. More research is needed to fully understand the impact of perceived difficulty for phonetics tasks on student learning and outcomes.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Nicole Whitworth ⓘ http://orcid.org/0000-0001-8349-5413

Data availability statement

The data that support the findings of this study are available from the author upon reasonable request.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.

Andrabi, M., Robinson, A., & Marques, F. (2022). Concept of perceived task difficulty: A systematic review. *Journal of Community & Public Health Nursing, 8*(9). https://doi.org/10.4172/2471-9846.1000363

Ashby, P. (2002). *Practical phonetics training and the nature of phonetic judgements* [PhD thesis]. University College London.

Ashby, P. (2003). Learning cardinal vowels. In M. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Conference of Phonetic Sciences (ICPhS)*) (pp. 3089–3092). Universitat Autonoma de Barcelona.

Ball, M. J., Müller, N., Klopfenstein, M., & Rutter, B. (2010). My client is using non-English sounds! A tutorial in advanced phonetic transcription part II: Vowels and diacritics. *Contemporary Issues in Communication Science & Disorders, 38*, 103–110. https://doi.org/10.1044/cicsd_36_F_103

Ball, M., Müller, N., Klopfenstein, M., & Rutter, B. (2009). The importance of narrow phonetic transcription for highly unintelligible speech: Some examples. *Logopedics Phoniatrics Vocology, 34* (2), 84–90. https://doi.org/10.1080/14015430902913535

Brown, A. (2016). Item response models for forced-choice questionnaires: A common framework. *Psychometrika*, *81*(1), 135–160. https://doi.org/10.1007/s11336-014-9434-9

Catford, J. C. (2001). *A practical introduction to phonetics* [Non-fiction] (2nd ed.). Oxford University Press.

Chen, A., Li, W., Chen, L., Wei, J., & Fu, W. (2022, July 19–22). How to implement efficient blended learning: The effects of teacher support and task difficulty. In *2022 International Symposium on Educational Technology (ISET)*, Hong Kong (pp. 234–238).

Child Speech Disorder Research Network. (2017). Good practice guidelines for transcription of children's speech samples in clinical practice and research.

Diez, D. M., Barr, C. D., & Cetinkaya-Rundel, M. (2019). *OpenIntro statistics*. OpenIntro.

Dioubina, O. I., & Pfitzinger, H. R. (2002). An IPA vowel diagram approach to analysing L1 effects on vowel production and perception. In *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002)* (pp. 2265–2268).

Greenhoot, A. F., & Dowsett, C. J. (2012). Secondary data analysis: An important tool for addressing developmental questions. *Journal of Cognition and Development*, *13*(1), 2–18. https://doi.org/10.1080/15248372.2012.646613

Howard, S., & Heselwood, B. (2002). Learning and teaching phonetic transcription for clinical purposes. *Clinical Linguistics & Phonetics*, *16*(5), 371–401. https://doi.org/10.1080/02699200210135893

Howard, S., & Heselwood, B. (2013). The contribution of phonetics to the study of vowel development and disorders. *Handbook of Vowels and Vowel Disorders*, *2*, 61. https://doi.org/10.4324/9780203103890.ch3

Kent, R. D., & Miolo, G. (1995). Phonetic abilities in the first year of life. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 303–334). Wiley.

Knight, R.-A. (2010). Transcribing nonsense words: The effect of numbers of voices and repetitions [Article]. *Clinical Linguistics & Phonetics*, *24*(6), 473–484. https://doi.org/10.3109/02699200903491267

Knight, R.-A., Bandali, C., Woodhead, C., & Vansadia, P. (2018). Clinicians' views of the training, use and maintenance of phonetic transcription in speech and language therapy. *International Journal of Language & Communication Disorders*, *53*(4), 776–787. https://doi.org/10.1111/1460-6984.12381

Knight, R.-A., & Maguire, E. (2011). The relationship between short-term memory and the phonetic transcription accuracy of speech and language therapy students. In *Proceedings of the Phonetics Teaching and Learning Conference (PTLC)* (pp. 21–24).

Knight, R.-A., Setter, J., & Cornelius, P. (2014). Articulatory phonetics. In N. Whitworth & R.-A. Knight (Eds.), *Methods in teaching clinical phonetics and linguistics* (pp. 23–45). J&R Press Ltd.

Knight, R.-A., Setter, J., & Whitworth, N. (2021). Pedagogical approaches. In R.-A. Knight & J. Setter (Eds.), *The Cambridge handbook of phonetics* (pp. 503–526). Cambridge University Press.

Ladefoged, P. (2001). *A course in phonetics*. Harcourt College Publishers.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell.

Mackenzie-Beck, J. (2003). Is it possible to predict students' ability to develop skills in practical phonetics? In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2833–2836). Universitat Autnoma de Barcelona.

Moran, M. J., & Fitch, J. L. (2001). Phonological awareness skills of university students: Implications for teaching phonetics. *Contemporary Issues in Communication Science and Disorders*, *28*(Fall), 85–90. https://doi.org/10.1044/cicsd_28_F_85

Nelson, T. L., Mok, Z., & Ttofari Eecen, K. (2019). Use of transcription when assessing children's speech: Australian speech-language pathologists' practices, challenges, and facilitators. *Folia Phoniatrica et Logopaedica*, *72*(2), 131–142. https://doi.org/10.1159/000503131

Peng, S., Man, K., Veldkamp, B. P., Cai, Y., & Tu, D. (2023). A mixture model for random responding behavior in forced-choice noncognitive assessment: Implication and application in organizational research. *Organizational Research Methods*. https://doi.org/10.1177/10944281231181642

Pollock, K. E., & Berni, M. C. (2001). Transcription of vowels. *Topics in Language Disorders*, *21*(4), 22–40. https://doi.org/10.1097/00011363-200121040-00005

Posit Software. (2023). *RStudio*. https://posit.co/products/open-source/rstudio/

The R Foundation. (2023). *The R project for statistical computing*. https://www.r-project.org/

Robinson, G. C., Mahurin, S. L., Richards, K. L., & Justus, B. (2011). Predicting difficulties in learning phonetic transcription: Phonemic awareness screening for beginning speech-language pathology students. *Contemporary Issues in Communication Science & Disorders*, *38*(Spring), 87–95. https://doi.org/10.1044/cicsd_38_S_87

Schneider, S., Nebel, S., Meyer, S., & Rey, G. D. (2022). The interdependency of perceived task difficulty and the choice effect when learning with multimedia materials. *Journal of Educational Psychology*, *114*(3), 443. https://doi.org/10.1037/edu0000686

Shaw, Á., & Yanushevskaya, I. (2022). Students' views and experiences of the training and use of phonetic transcription in speech and language therapy – The Irish perspective. *Clinical Linguistics & Phonetics*, *36*(2–3), 276–291. https://doi.org/10.1080/02699206.2021.1874055

Stemberger, J. P., & Bernhardt, B. M. (2019). Phonetic transcription for speech-language pathology in the 21st Century. *Folia Phoniatrica et Logopaedica*, *72*(2), 75–83. https://doi.org/10.1159/000500701

Stephanou, G., Kariotoglou, P., & Dinas, K. (2011). University students' emotions in lectures: The effect of competence beliefs, value beliefs and perceived task-difficulty, and the impact on academic performance. *International Journal of Learning*, *18*(1), 45–72. https://doi.org/10.18848/1447-9494/CGP/v18i01/47453

Street, K. E., Stylianides, G. J., & Malmberg, L.-E. (2022). Differential relationships between mathematics self-efficacy and national test performance according to perceived task difficulty. *Assessment in Education Principles, Policy & Practice*, *29*(3), 288–309. https://doi.org/10.1080/0969594X.2022.2095980

Teoh, A. P., & Chin, S. B. (2009). Transcribing the speech of children with cochlear implants: Clinical application of narrow phonetic transcriptions [Article]. *American Journal of Speech-Language Pathology*, *18*(4), 388–401. https://doi.org/10.1044/1058-0360(2009/08-0076)

Titterington, J., & Bates, S. (2018). Practice makes perfect? The pedagogic value of online independent phonetic transcription practice for speech and language therapy students. *Clinical Linguistics & Phonetics*, *32*(3), 249–266. https://doi.org/10.1080/02699206.2017.1350882

Titterington, J., & Bates, S. (2021). Teaching and learning clinical phonetic transcription. In M. J. Ball (Ed.), *Manual of clinical phonetics* (pp. 175–186). Routledge.

White, S., Hurren, A., James, S., & Knight, R. A. (2022). 'I think that's what I heard? I'm not sure': Speech and language therapists' views of, and practices in, phonetic transcription. *International Journal of Language & Communication Disorders*, *57*(5), 1071–1084. https://doi.org/10.1111/1460-6984.12740

Whitworth, N. (2008). *Hard sounds & easy sounds: SLT students' perceptions of IPA sounds*. Fifth Colloquium of the British Association of Clinical Linguists (BACL), University of Reading.

Whitworth, N. (2011). *SLT students' perception of IPA sounds: An update on vowels*. Third Colloquium of the British Association of Clinical Linguists (BACL), Leeds Metropolitan University.

Wikström, J., & Setter, J. (2011). Speech and language therapy (SLT) students' production and perception of cardinal vowels: A longitudinal case study of six speech and language therapy students. *Leeds Working Papers in Linguistics and Phonetics*, *16*, 51–82. https://www.latl.leeds.ac.uk/wp-content/uploads/sites/49/2019/05/Wikstrom-Setter_2011.pdf

Yücel, A. G. (2022). *Task Complexity and working memory in performing listen-to-speak integrated tasks in a second language* [MA thesis]. Boğaziçi University.