2025. Proc Ling Soc Amer 10(1). 5932. https://doi.org/10.3765/plsa.v10i1.5932.

Quantifying metalinguistic awareness of sociophonetic features

Lisa Sullivan*

Abstract. Metalinguistic awareness of sociophonetic features may vary based on social or individual factors (e.g. dialect region, production, perception), as well as properties of the dialects or features themselves (e.g. their markedness). It is necessary to quantify metalinguistic awareness in order to consider these relationships statistically. This study tests a method of quantifying metalinguistic awareness using three tasks (written dialect description, written dialect identification, auditory dialect identification) and four sociophonetic features of North American English (/æg/-raising, /aj/-monophthongization and, Canadian raising of /aj/ and /aw/, considered separately). It finds that it is possible to quantify metalinguistic awareness, but that other modes of folk linguistic awareness, such as detail and accuracy, contribute differentially to the tasks used in this study.

Keywords. metalinguistic awareness; folk linguistic awareness; sociophonetics; North American English

1. Introduction. Research on metalinguistic awareness (MLA) of sociophonetic features suggests that differences in MLA may be associated with the markedness of dialects or features, and individual participants' production, perception or dialect region (Johnstone & Kiesling (2008); Ruch (2018); Sullivan (2022)). To consider these relationships statistically, it is necessary to quantify MLA. The current study attempts to quantify this using three tasks designed to tap into different levels of MLA and four sociophonetic features hypothesized to have different levels of MLA: /æg/-raising (BAG), /aj/-monophthongization (AM) and, Canadian raising of /aj/ (FCR) and /aw/ (BCR). BAG and FCR are expected to have lower MLA than BCR, which is highly, but inaccurately stereotyped, and AM, a feature of the Southern US dialect.

I begin in section 1 with a discussion of metalinguistic awareness and its relation to folk linguistic awareness (Preston 1996), a brief description of the features under investigation and the study goals and hypotheses. Sections 2 and 3 outline the methods and analysis, respectively. These are followed by the results in section 4 and a discussion of the findings in section 5.

1.1. Metalinguistic Awareness & Folk Linguistic Awareness. For the purposes of this study I define metalinguistic awareness to be the degree of awareness people have of phonetic variants (Sullivan (2022)). For example, how aware are you of the difference between the raised variant of /æg/ used by speakers who /æg/-raise and the unraised variant used by people who don't. This metalinguistic awareness can be thought of as having three levels which exist on a continuum. At the lowest level, no awareness, people are unaware of phonetic variants (e.g. they don't recognize a difference between raised and unraised /æg/). In the middle, phonetic awareness occurs when people are aware of a difference between variants, but it has no social meaning (e.g. they recognize a difference between raised and unraised /æg/, but have no social meaning associated with it). At the highest level, people are aware of phonetic variants and associate social meaning with those variants (e.g. they associate people with the raised variant of /æg/

* Thank you to Jessamyn Schertz, Nathan Sanders, Derek Denis, Yoonjung Kang, Jack Chambers, Valerie Fridland, Valerie Freeman and Beth MacLeod for their feedback on this project. This project is supported in part by funding from the Social Sciences and Humanities Reserach Council of Canada. Author: Lisa Sullivan, Oklahoma State University (lisa.sullivan10@okstate.edu).

with a certain geographic region).

Previous research on metalinguistic awareness has shown that the degree of metalinguistic awareness someone has of a feature may be related to that person's use of that feature, the markedness of the feature and/or dialect it is a feature of, and the dialect region the person is from. In one study, Johnstone & Kiesling (2008) examined the perception of monophthongal /aw/ in Pittsburgh using a matched guise task and interviewed individual participants about their metalinguistic awareness. They found that participants who did not use monophthongal /aw/ were more aware of it than participants who did use it, suggesting that using a feature makes people less likely to have metalinguistsic awareness of it.

In another, Ruch (2018) ran an open dialect description task for two Swiss German dialects (Zurich and Grison). She found that people described the more marked Grison dialect in more detail than the less marked Zurich dialect, and that they described more marked dialect features more than less marked features, suggesting that markedness of features and dialects leads to increased metalinguistic awareness. Finally, in a third study, Sullivan (2022) examined the awareness of /æg/-raising across North America using dialect description and identification tasks. She found that participants from dialect regions without /æg/-raising have higher metalinguistic awareness of /æg/-raising than those from regions where /æg/-raising is present. This suggests that the region a person grew up in may influence their metalinguistic awareness.

Metalinguistsic awareness is related to Preston (1996)'s four modes of folk linguistic awareness: (1) availability (are speakers aware of variants), (2) accuracy (are they able to describe variants accurately), (3) detail (how specific and detailed is their awareness) and (4) control (are speakers able to perform the variants they are describing). Metalinguistic awareness is particularly associated with the availability mode as both have to do with people's awareness of variants existing, though other modes may also contribute to metalinguistic awareness (at least in the tasks used in this study).

Preston describes a 4-level scale for the availability mode ranging from common (something people readily discuss without prompting) through suggestible (requires some prompting) and available (requires significant prompting, but does social work) to unavailable (does not do social work). The relationship between the levels of metalinguistic awareness and the levels of availability is shown in Figure 1. The unavailable level of availability spans the no awareness and phonetic awareness levels of MLA as it is possible to be able to notice a difference, but not to assign it any social meaning. The available level spans the higher end of phonetic awareness and lower end of social awareness, as higher levels of phonetic awareness could lead to inferences about what social work they may do. The suggestible and common levels align with higher levels of social awareness, with common corresponding with the highest levels of social awareness.

1.2. SOCIOPHONETIC FEATURES. To test the plausibility of quantifying metalinguistic awareness, four sociophonetic features of North American English with varying geographic distributions and levels of folk linguistic awareness will be used: /æg/-raising, Canadian Raising (Chambers 1973) of /aj/ and /aw/ (hereafter /aj/-raisng and /aw/-raising), and /aj/-monophthongization.

/æg/-raising is the process whereby /æ/ is raised before /g/ in some dialects of North American English. Acoustic documentation has found /æg/-raising in Canada as well as the Upper Midwest, Pacific Northwest and California in the United States (see Sullivan 2022 for a summary). Similar to the acoustic studies, Stanley (2022) found that people from Canada and Northern portions of the United States self-reported more /æg/-raising than those further south in the US. Met-

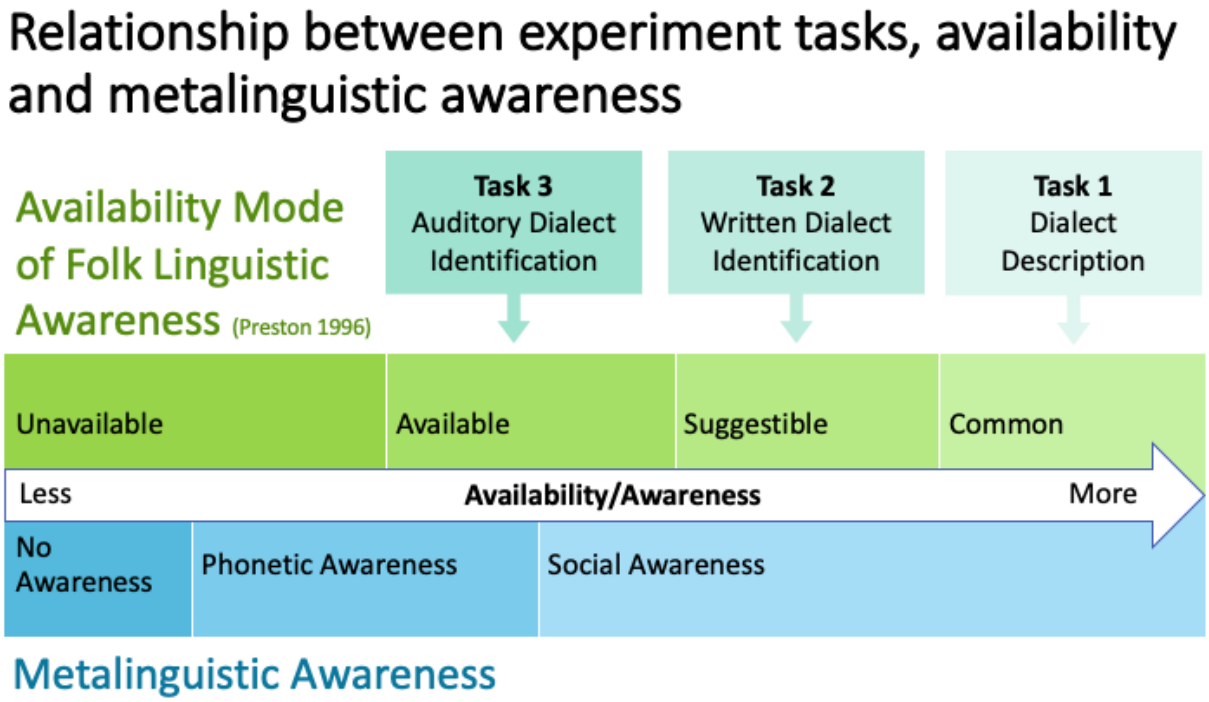

Figure 1. Relationship between metalinguistic awareness, the availability mode of folk linguistic awareness and the tasks in this study.

alinguistic awareness of /æg/ raising may be relatively low as this feature hasn't been stereotyped, though Stanley (2018) and Swan (2016) find some level of social awareness of /æg/-raising as a regional (and rural) feature in the Pacific Northwest and Vancouver, and Sullivan (2022) reports anecdotes of Americans living in Toronto, Ontario, Canada having social awareness of the feature as a marker of Canadian English, but Canadians begin unaware of it, suggesting that metalinguistic awareness varies based on social factors (e.g. rurality, geographic region).

Canadian Raising involves the raising of /aj/ and /aw/ to [ʌj] and [ʌw], respectively, before voiceless obstruents. I will consider these variants as separate features because of their different geographic distributions. While both are widespread in Canada, /aj/-raising has a much wider distribution across the United States than /aw/-raising which is limited to some northern states (e.g. Minnesota, Wisconsin) (see Onosson 2018 for further discussion). Furthermore, /aw/-raising is highly stereotyped as a feature of Canadian English while /aj/-raising is not. The stereotype of /aw/-raising as [uw], as in the phrase *'oot and aboot*, while well known, is also inaccurate, which could lead to easy description, but difficulty with hearing this feature. In contrast, /aj/-monophthongization is a highly salient feature of Southern American English which is frequently performed in imitations of an American accent (Plichta & Preston 2005). While this feature is readily performed, it is not clear how easy it is to describe, which could lead to a pattern where it is easily identified, but not described.

1.3. STUDY GOALS AND HYPOTHESES. The primary goal of this study is to test a methodology for quantifying folk linguistic awareness which involves three tasks (1) a written dialect description task (Task 1) in which participants are given a region and asked to describe accent features of that region, (2) a written dialect identification task (Task 2) in which participants are

given a word and asked to describe variations of that word and where they are used, and (3) an auditory dialect identification ask (Task 3) in which participants hear a speaker saying a word and state where they think the speaker is from. While all three tasks are expected to require at least some level of social awareness (available, suggestible or common availability on the folk linguistic awareness scale), the dialect description task is expected to require the highest level of social awareness, and the auditory dialect identification task, the lowest. The hypothesized relationship between the tasks and metalinguistic awareness and availability is illustrated in Figure 1.

In addition to differences across tasks, differences are expected across the four sociophonetic features under examination. Higher metalinguistic awareness is expected for /aw/-raising and /aj/-monophthongization than the other features since folk linguistic awareness of these features is high and they are a common part of lay discourse on North American accents. It is also possible that people may have higher metalinguistic awareness of /æg/-raising than /aj/-raising since it appears to be more geographically widespread, and there does appear to be some metalinguistic awareness of it, though only by select groups of speakers.

A secondary goal is to test if and how the detail and accuracy modes of folk linguistic awareness might impact results on these tasks. Features that have high accuracy levels may be more readily identified in the auditory dialect identification task than those with lower accuracy levels. On the other hand, features with high detail levels may be more easy for participants to describe in the two written tasks than those with low detail levels. These modes may be particularly relevant for features like /aj/-monophthongization, which has high accuracy and an unclear detail level, and /aw/-raising, which has low accuracy but high detail. It could be that participants display higher scores on the written tasks (Tasks 1 and 2) than the auditory task (Task 3) for /aw/-raising while they show higher scores on the auditory task than the written ones for /aj/-monophthongization.

2. Method. Participants completed a written dialect description task, written dialect identification task and auditory dialect identification task. Complete details of the methodology can be found in Sullivan (2022).

2.1. PARTICIPANTS. 61 native speakers of North American English from across the US and Canada (33 female, 28 male) completed the study. The age range of the participants was 18 to 72, with a mean age of 35.7. Participants were recruited on prolific.io. The distribution of participants across the US and Canada can be seen in Figure 2.

2.2. TASK 1: WRITTEN DIALECT DESCRIPTION. Participants were asked to describe the accent of speakers from Alabama, Boston, Canada, Minnesota, Newfoundland and Labrador, the Ottawa Valley, and Seattle, including words or sounds people from these regions said differently than others. These regions were chosen based on a pre-test and to include regions where the features under investigation would be present.

2.3. TASK 2: WRITTEN DIALECT IDENTIFICATION. Participants were asked to describe different pronunciations of 8 monosyllabic words, including *about* (/aw/-raising), *bag* (/æg/-raising) and *right* (/aj/-rasing, /aj/-monophthongization), as well as which speakers use the different pronunciations. These words were chosen to include the features under investigation.

2.4. TASK 3: AUDITORY DIALECT IDENTIFICATION. Stimuli for this task were recorded by 9 native speakers of North American English (1 recording in 2 accents), including 6 Canadians from the Ottawa Valley (3 male, 3 female) and 3 Americans from the Midwest, California and

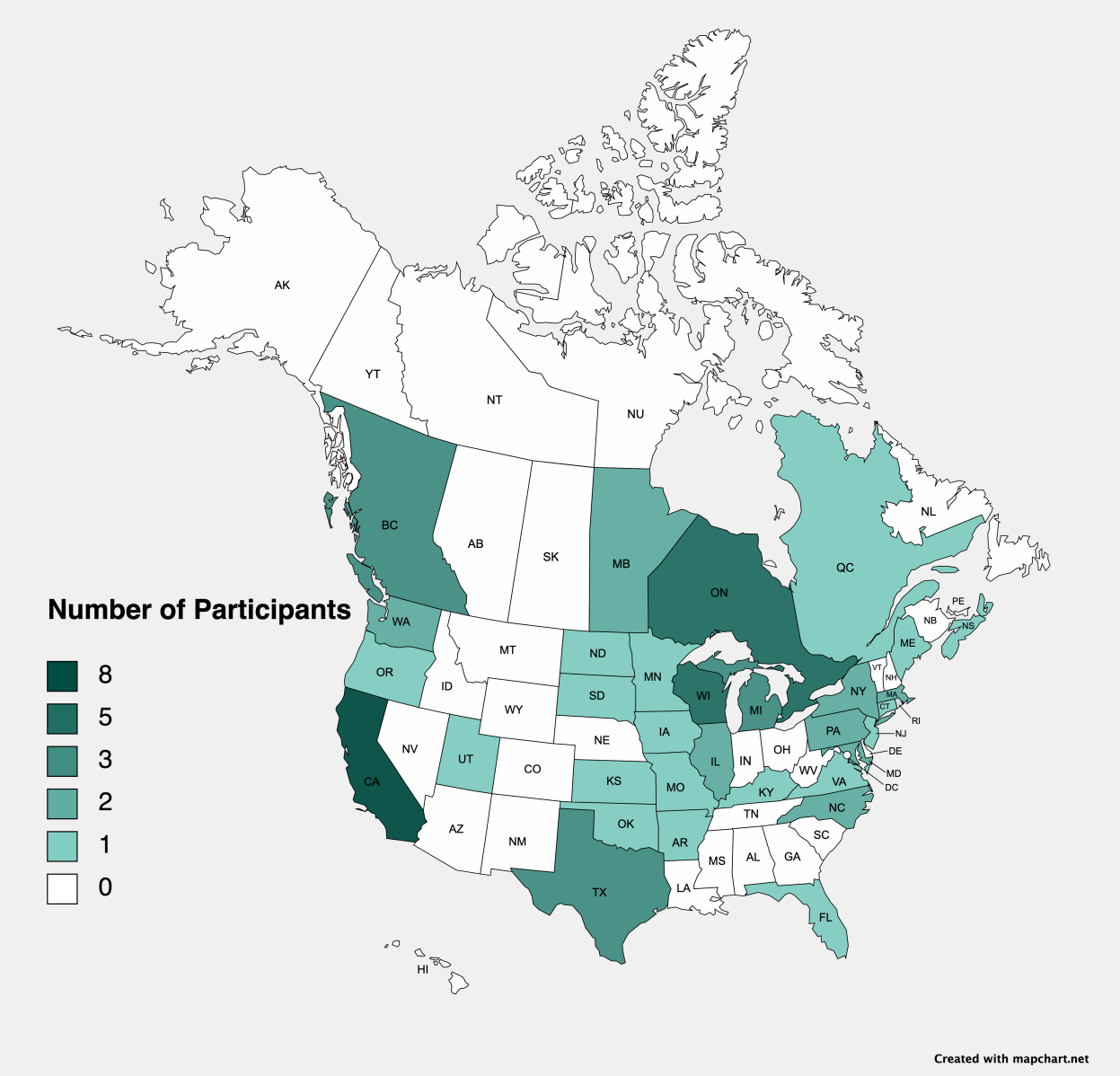

Figure 2. Regional distribution of participants in the metalinguistic awareness survey

the US South (all male). Speakers read a 30 word list twice. For the raising features, the list included 2 minimal pairs containing the vowel in a raising and non-raising context (e.g. *bag*/*back* and *lag*/*lack* for /æg/-raising features). For /aj/-monophthongization, the list included 2 words (*tight* and *bite*).

Praat (Boersma & Weenink 2021)'s change gender feature was used to make the female speakers' tokens sound male to avoid possible speaker gender effects, and to make the male speaker who recorded 2 accents (Southern/General American) tokens sound more distinct. On each trial, participants heard one of the speakers saying a monosyllabic word and had to guess where they were from. Target words were *bag* (/æg/-raising), *tight* (/aj/-raising, /aj/-monophthongization)

and *lout* (/aw/-raising). Participants heard 10 tokens from each speaker, including each target word and 7 filler words, for a total of 100 trials. Trials were randomized by participant.

2.5. OVERALL PROCEDURE. The procedure was conducted online using jsPsych (de Leeuw 2015). Participants completed the tasks in order followed by a language background and demographic questionnaire.

3. Analysis.

3.1. SPEAKER ANALYSIS. To determine if speakers participated in each of the processes under investigation, words containing the target features were segmented in Praat and F1 and F2 at 13 equidistant points across the trajectory (including the onset and offset) were extracted using by-speaker formant settings. Formant measures were converted to Bark using Zwicker & Terhardt (1980)'s formula.

For the three raising features, average F1 values were calculated for each speaker's raising and non-raising environment by averaging the 7 measured points in the middle 50% of the vowel for all four tokens in each context (2 words/context*2 repetitions). This was done to minimize the effects of variation in measurement at individual points across the vowels. A measure of raising was then calculated by each speaker by subtracting their average F1 in the non-raising context from their average F1 in the raising context.

Figures 3 – 5 show the degree of raising by speaker, the black line represents a difference of 1Bk between the raising and non-raising contexts. For /aj/-raising and /æg/-raising, speakers with a difference higher than 1 were categorized as having that feature for the purposes of the analysis. For /aj/-raising, this included 5 Canadians, the American from the Midwest and the American speaking in a general American accent. For /æg/-raising this included all 6 Canadians. For /aw/-raising, there was a clear distinction between the American speakers, who had negative or near 0 differences, and the Canadian speakers, so the Canadians were all categorized as having raising.

For /aj/-monophonogization, the vowel trajectories exculding the end-point values (11 points) were plotted for F1 and F2 (Figure 6). A clear difference is noticeable between the speaker from the US south, who has a much more monophthongal trajectory, particularly for F2, compared to the others. This speaker was categorized as participating in /aj/-monophthongization.

3.2. PARTICIPANT RESPONSE CODING. Participants received a score of 1 for each feature in Task 1 if they described a feature in a reasonably correct way in one of the regions where it was known to occur, and a 0 otherwise. For /aj/-raising and /aw/-raising, this included Canada (or regions therein) and Minnesota. For /æg/-raising, this included the Canada (or regions therein), Minnesota and Seattle. For /aj/-monophongization, this included Alabama.

In Task 2, participants received a score of 1 if they described a feature in a reasonably correct way and correctly identified a region it was known to occur in, and a 0 otherwise. This includes the regions mentioned above in addition to the Pacific Northwest and Upper Midwest for /æg/-raising and the US South or any southern state for /aj/-monophthongization.

In Task 3, participants were scored by word that they heard. They received a score of 1 for a particular feature if they identified the speaker as being from a region with that feature and 0 otherwise. For each feature, a participant's scores for the speakers identified as having that feature were averaged to obtain an individual score. (e.g. for /æg/-raising, the scores the participant received for the 6 Canadian speakers were averaged to obtain a single score).

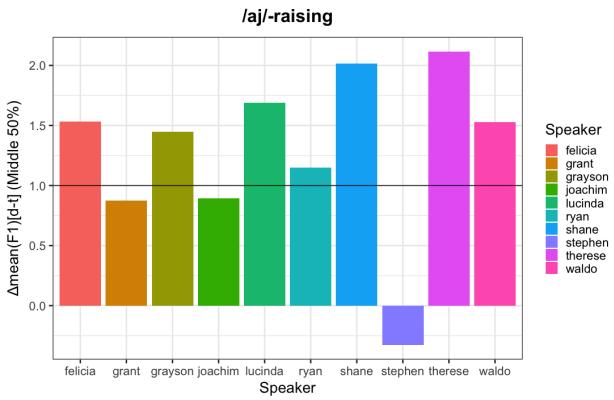

Figure 3. Degree of /aj/-raising by speaker

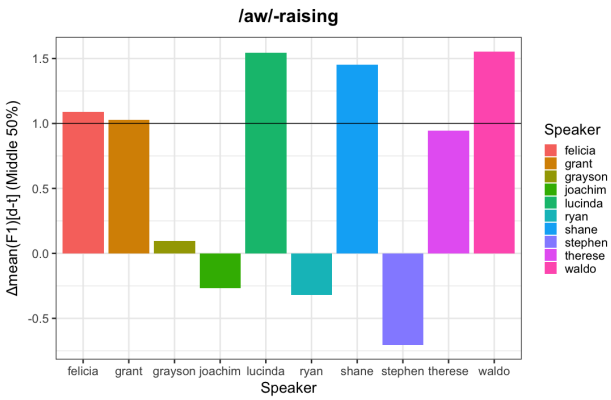

Figure 4. Degree of /aw/-raising by speaker

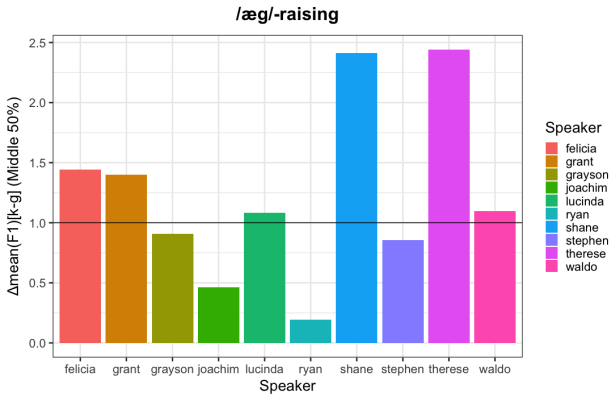

Figure 5. Degree of /æg/-raising by speaker

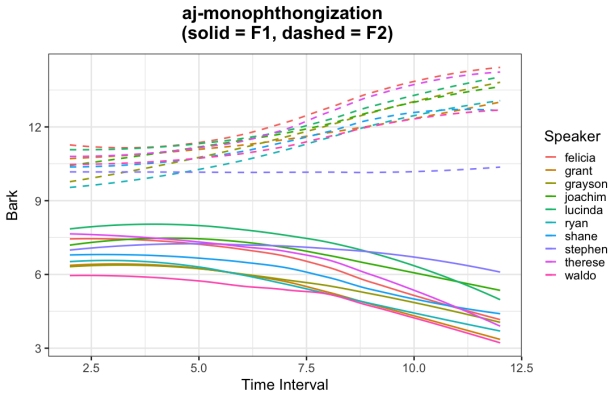

Figure 6. Degree of /aw/-monophonogization by speaker

3.3. STATISTICAL ANALYSIS. As this analysis is exploratory in nature, a series regression analyses, conditional inference trees and random forests analyses were conducted. Statistical analyses were conducted in R (R Core Team 2020) using base functions as well as functions from the lme4 (Bates et al. 2015), buildmer (Voeten 2023) and party (Hothorn et al. 2006) packages.

To compare tasks, a mixed effects linear regression model was constructed using *lmer()* with score as the predictor variable and task as the response variable. Random intercepts were included for feature and participant. Task was coded ordinally for comparisons between Tasks 1 and 2 and between Tasks 2 and 3. Year of Birth was z-score normalized using *scale()*. Looking at features individually, full interaction step-wise mixed effects linear regression models were constructed using *bulidmer()*. The response variable for each model was the score, and the predictor variables were task, participant year of birth and participant gender. A random intercept was included for participant. Gender was simple coded (female = -0.5, male = 0.5), and task was coded as described above.

Conditional inference tree and random forest analyses were constructed using the *ctree()* and *cforest()* functions, respectively. Complete dataset analyses constructed with score as the response variable and task, feature, year of birth and gender as response variables. By-feature analyses were constructed with task, year or birth and gender as response variables.

Finally, to test that participants were responding as expected in Task 3 (Auditory Dialect

Identification) for the raising features, simple linear regression models were constructed with mean score each speaker receive as the response variable and change in F1 as the predictor variable.

4. Results. The overall results can be found in Figure 7, which shows the mean score for each task (left) and feature (right). Looking at the tasks, Task 1 (Written Dialect Description) has the lowest score, and Task 3 (Auditory Dialect Identification), the highest. The regression model shows a significant difference between Task 1 and Task 2 (Written Dialect Identification) (Est. = -0.250, *SE* = 0.032, df = 666, t = -7.73, p ¡ 0.001), but not between Tasks 2 and 3 (Est. = -0.037, *SE* = 0.032, df = 666, t = -1.17, p = 0.244). The feature plot shows that different features have different levels of metalinguistic awreness, with /aw/-raising (BCR) having the highest MLA scores, followed by /aj/-monophonogization (AM), then /æg/-raising (BAG), and finally /aj/-raising (FCR) with the lowest score.

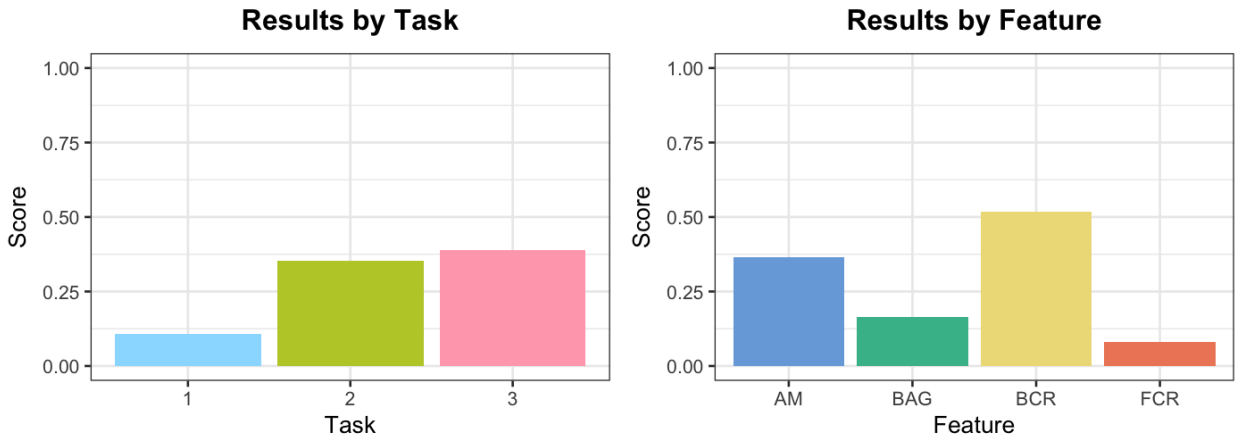

Figure 7. MLA Scores by Task (left) and by Feature (right)

The results of the conditional inference tree and random forest analyses are shown in Figure 8. Like the mean values in Figure 7 and the regression model for task, they suggest that task and feature, but neither gender nor year of birth, are important factors for predicting MLA scores. The conditional inference tree also reveals that different features display different patterns across the three tasks, though exactly how these break down is not clear from the tree. Figure 9 shows this breakdown and suggests that different features display different patterns across the tasks, with /æg/-raising and /aw/-raising showing the highest MLA scores in Task 2 (Written Identification) while /aj/-monophthongization and /aj/-raising show the highest scores in Task 3 (Auditory Identification). To explore this, each feature will be discussed individually in the subsequent sections.

4.1. /AJ/-RAISING. Results for /aj/-raising show low metalinguistic awareness in all tasks, including none in Task 1, and more awareness in Task 2 than Task 3. The regression analysis (Table 1) shows a significant difference between Task 1 and Task 2 whereby participants have higher scores in Task 2 and an interaction between gender and Tasks 2 and 3 whereby male participants have similar scores in each task but female participants have higher scores in Task 3 than Task 2.

The conditional inference tree and random forest analyses (Figure 10) reveal that Task is the most important factor for this feature, with a split between lower scores in Task 1 and higher

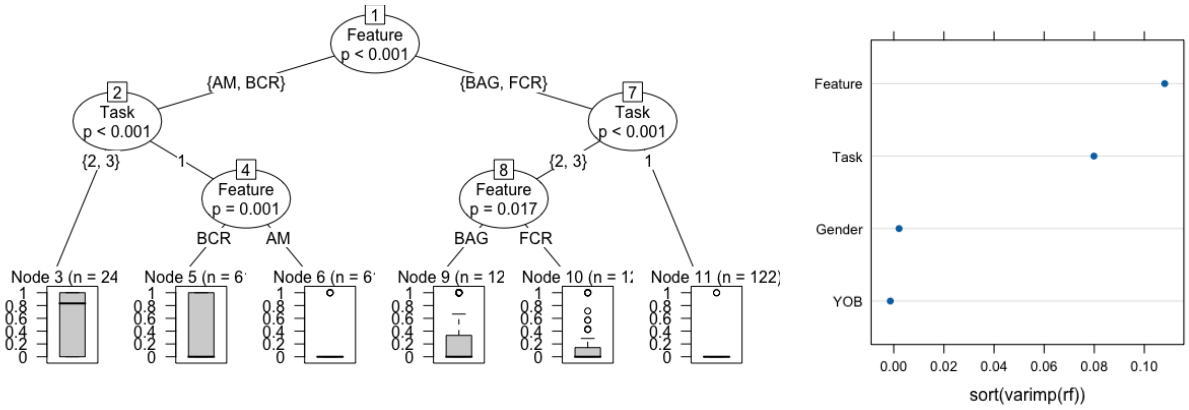

Figure 8. Conditional inference tree and random forest for the full model

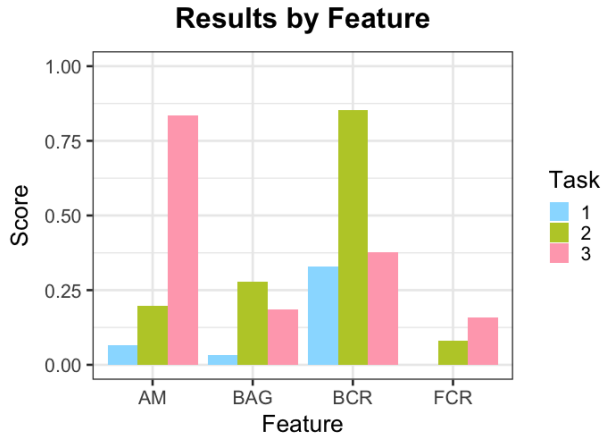

Figure 9. MLA Scores by Task and by Feature

	Est.	*SE*	t-value	p-value	
(Intercept)	0.0793	0.0146	5.45	<0.001	***
Task 1 vs Task 2	-0.0839	0.0357	-2.35	0.0197	*
Task 2 vs Task 3	-0.0703	0.0357	-1.97	0.0503	.
Gender: F vs M	-0.0260	0.0291	-0.894	0.373	
Task:1 vs 2 * Gender	-0.0465	0.0713	-0.653	0.515	
Task: 2 vs 3 * Gender	0.171	0.0713	2.40	0.0174	*

Table 1. Results of the step-wise mixed effects linear regression model for /aj/-raising. Model: MLA $\sim$ Task*Gender

scores in Tasks 2 and 3. The random forest also shows that participant gender may have a larger effect for this feature than for the others.

4.2. /ÆG/-RAISING. Results for /æg/-raising show fairly low metalinguistic awareness overall, with Task 1 scores being much lower than Task 2 and 3 scores, and Task 3 scores being lower than Task 2 scores. The regression model (Table 2) for this feature confirms this, finding a sig-

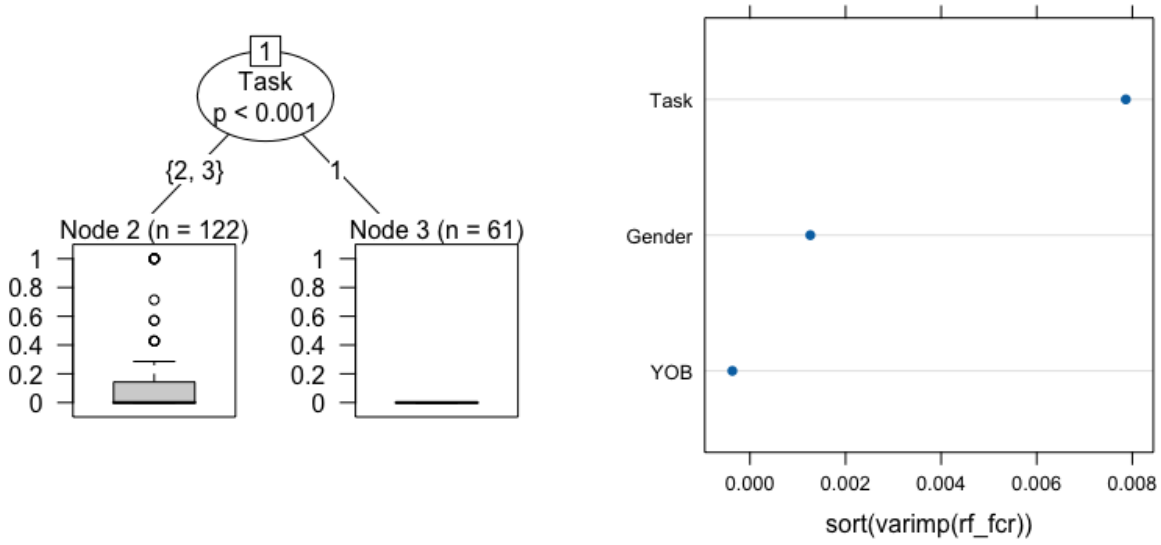

Figure 10. Conditional inference tree and random forest for /aj/-raising

nificant difference between Task 1 and Task 2, but not between Tasks 2 and 3. The conditional inference tree (Figure 11) likewise shows a split between Task 1 and the other two tasks while the random forest (Figure 11) shows that Task is the most important factor for determining metalinguistic awareness of this feature.

	Est.	*SE*	t-value	p-value	
(Intercept)	0.16576	0.02229	7.436	<0.001	***
Task 1 vs Task 2	-0.24590	0.05460	-4.503	<0.001	***
Task 2 vs Task 3	0.09290	0.05460	1.701	0.0906	.

Table 2. Results of the step-wise mixed effects linear regression model for /æg/-raising. Model: MLA $\sim$ Task

4.3. /AJ/-MONOPHONOGIZATION. Results for /aj/-monophthongization show the expected pattern. Participants had the highest metalinguistic awareness scores in Task 3 and the lowest in Task 2. The regression analysis confirms this, finding significant effects between Task 1 and Task 2, and between Task 2 and Task 3. Figure 9 also shows that awareness in Task 3 appears to be much higher than in the other two tasks. This is confirmed by the conditional inference tree (Figure 12) which splits the data between task 3 and the other two tasks. The random forest (Figure 12) shows that Task is the most important factor for determining metalinguistic awareness scores for this feature.

4.4. /AW/-RAISING. The pattern for /aw/-raising in Figure 9 shows similar scores for Tasks 1 and 3, with Task 3 having a slightly higher score. The scores for Task 2 were highest by a large margin for this feature. The regression model confirms this, showing significant differences between Task 2 and each of the other two tasks. The conditional inference tree (Figure 13) likewise

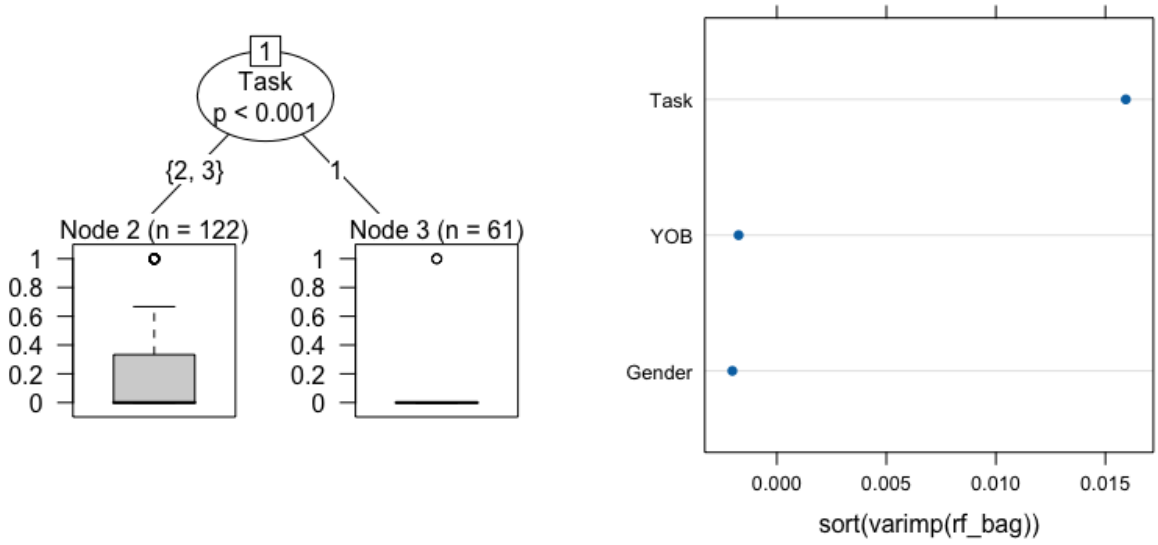

Figure 11. Conditional inference tree and random forest for /æg/-raising

	Est.	*SE*	t-value	p-value	
(Intercept)	0.36612	0.02569	14.252	<0.001	***
Task 1 vs Task 2	-0.131	0.0629	-2.08	0.0386	*
Task 2 vs Task 3	-0.63934	0.06292	-10.160	<0.001	***

Table 3. Results of the step-wise mixed effects linear regression model for /aj/-monophthongization. Model: MLA ∼ Task

shows a split between Task 2 and the other two tasks. The random forest, as with the other features, shows that Task is the most important factor determining metalinguistic awareness score.

	Est.	*SE*	t-value	p-value	
(Intercept)	0.519	0.0283	18.3	<0.001	***
Task 1 vs Task 2	-0.525	0.0694	-7.55	<0.001	***
Task 2 vs Task 3	0.475	0.0694	6.85	<0.001	***

Table 4. Results of the step-wise mixed effects linear regression model for /aw/-raising. Model: MLA ∼ Task

4.5. VERIFICATION OF TASK 3 (AUDITORY DIALECT IDENTIFICATION). Figure 14 shows the correlation between speakers' degree of raising in production and the average MLA score they received in Task 3, with higher scores showing more awareness. The line in each plot is the line of best fit, while each dot represents one speaker. There appears to be a positive correlation between speakers' degree of raising in production and their mean MLA score for all three features. The simple linear regression models (Table 5) show that this correlation is significant for /æg/-raising and /aw/-raising, but not for /aj/-monophthongization. This latter non-significant finding

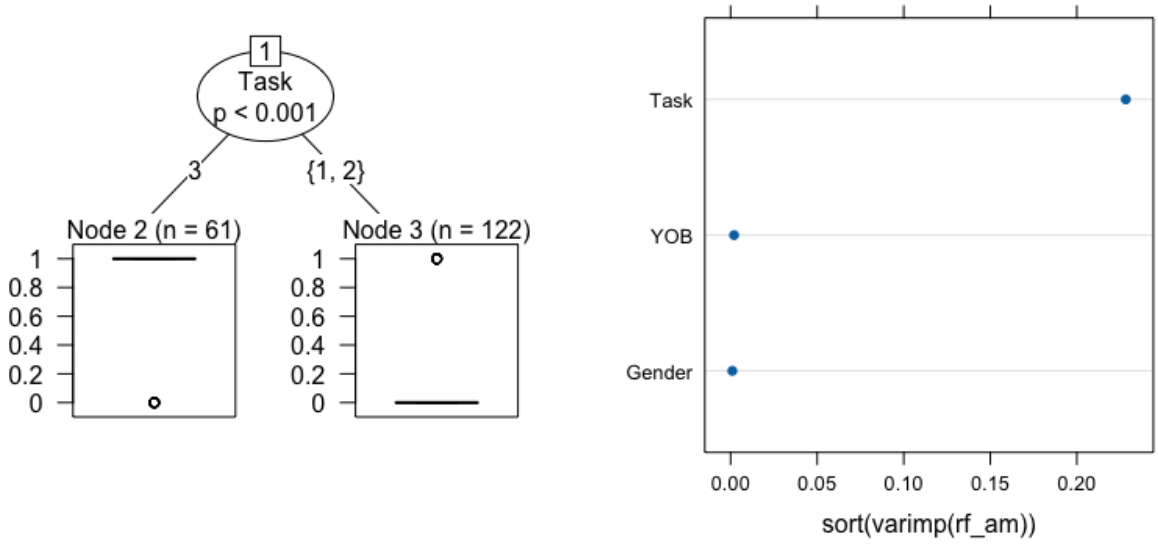

Figure 12. Conditional inference tree and random forest for /aj/-monophthongization

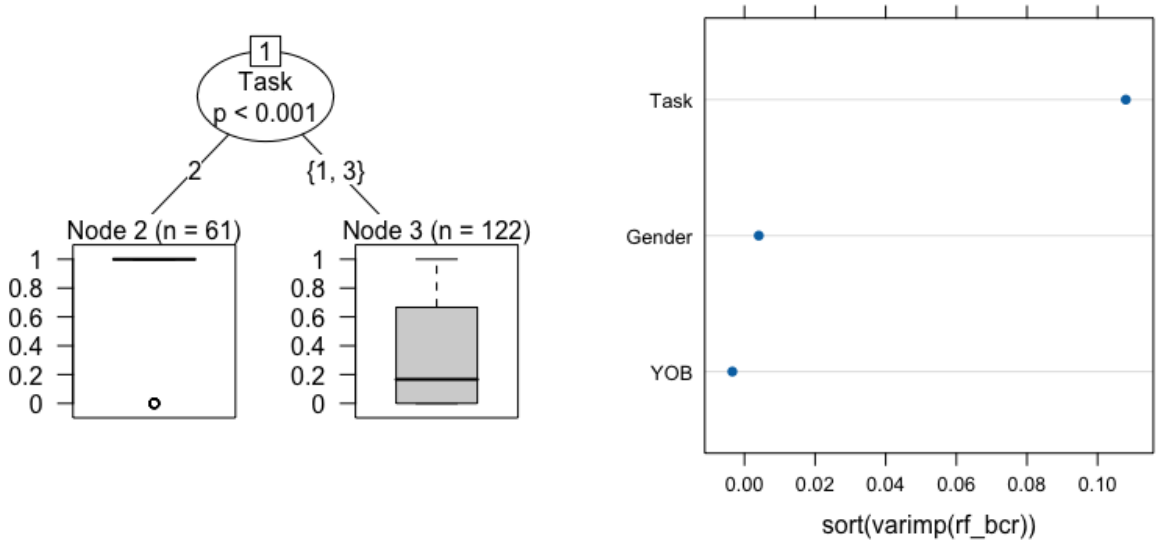

Figure 13. Conditional inference tree and random forest for /aw/-raising

is perhaps not unexpected given the generally low level of MLA for this feature, and fact that most speakers have fairly similar levels of raising for this feature. These findings confirm that the participants were behaving as expected in this task.

4.6. SUMMARY. In summary, we find a significant difference between Task 1 and Tasks 2 and 3, and that phonetic feature and task are the largest contributors in determining metalinguistic awareness. Looking at each feature individually, we see that /aw/-raising and /aj/-monophthongization

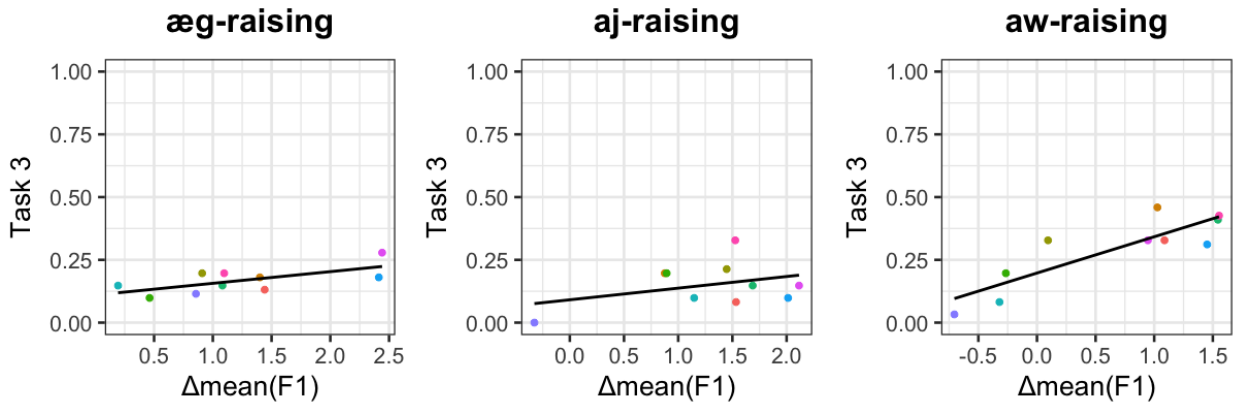

Figure 14. Correlation between speaker's mean degree of raising and their mean score on task 3 for /æg/-raising, /aj/-raising and /aw/-raising

Feature	Est.	*SE*	t-value	p-value	
/æg/-raising	0.0466	0.0186	2.51	<0.0365	*
/aj/-raising	0.0464	0.0419	1.11	<0.301	
/aw/-raising	0.144	0.0297	4.86	<0.00126	**

Table 5. Results of the simple linear regression models for the three raising features

have higher levels of metalinguistic awareness than /æg/-raising and /aj/-raising. Comparing /aw/-raising and /aj/-monophthongization shows that these two features have different patterns. Task 2 yields the highest MLA scores for /aw/-raising while Task 3 yields the highest scores for /aj/-monophthongization. Verification of Task 3 shows that participants are responding to this task as expected.

5. Conclusion. The primary goal of this study was to attempt to quantify metalinguistic awareness through a series of dialect description and identification tasks. The results suggest that this is possible, but that the relationship between the tasks may be more complicated than originally hypothesized (see Figure 1).

Looking first at the sociophonetic features tested in the study, participants showed higher metalinguistic awareness of /aw/-raising and /aj/-monophthongization than /æg/-raising and /aj/-raising, which is expected given that speakers are known to have high metalinguistic awareness of these features. Participants also seemed to have higher metalinguistic awareness of /æg/-raising than /aj/-raising. This is also not unexpected given the more limited geographic distribution of /æg/-raising than /aj/-raising, and the evidence of metalinguistic awareness in select groups of speakers. Notably, the only two participants who identified /æg/-raising in Task 1 were an American living in Toronto and someone from Vancouver, which aligns with the previous findings that these groups of speakers have metalinguistic awareness of /æg/-raising (Swan 2016; Stanley 2018; Sullivan 2022).

Turning to differences between tasks, participants had higher scores in the two identification tasks (Tasks 2 and 3) than in the description task (Task 1), indicating that higher levels of metalinguistic awareness were required for Task 1. However, the difference between the identification tasks was not significant, suggesting that the auditory identification task (Task 3) did not

require higher levels of metalinguistic awareness than the written identification task (Task 2).

This latter finding may be attributable in part to the differences in the way scores were calculated between the two tasks. In Task 2, participants received a single score of 0 or 1 for each feature based on their response for one word while, their scores of 0 or 1 were averaged across speakers for Task 3[1]. This method resulted in lower scores in Task 3 by participant as it was rare for a participant to correctly identify the region for all the speakers who participated in a particular sociophonetic process. Another possibility is that, since participants only heard each speaker say each word once before guessing the region, that this made the task more difficult than if they had been able to hear a series of words.

A final reason for this lack of difference relates to this study's secondary goal of testing the contributions of accuracy and detail to the tasks used to quantify metalinguistic awareness. It may be that these other aspects of folk metalinguistic awareness contributed to participant scores in the two identification tasks. Since Task 2 required participants to be able to describe what the features sounded like, it would have required some level of detail. On the other hand, since Task 3 required categorization based on perception, it would have required a certain level of accuracy. While /aw/-raising and /aj/-monophthongization both have high metalinguistic awareness, their accuracy and detail are different in ways that align with these differences between tasks.

In the case of /aw/-raising, it is highly but inaccurately stereotyped as *'oot and aboot* (for *out and about*), which lends itself to easy description, but may lead to difficulties when people hear isolated words, as observed in the higher metalinguistic awareness scores participants achieved in Task 2 relative to Task 3[2]. Notably, in participant responses in the written tasks, there were two kinds of descriptions of /aw/-raising. Participants from /aw/-raising regions or with exposure to /aw/-raising often noted the stereotype, but also that it is not quite accurate, whereas other participants just noted the stereotype, which suggests differences in degree of accuracy between these two groups. It is not surprising, then, that some participants may be able to describe the feature in the written task, but then not be able to hear it in the auditory task.

It is also worth mentioning that the word used for /aw/-monophthongization in Task 2 was *about* which is part of this stereotype, and may have contributed to the high Task 3 scores in a way a less stereotyped word like *shout* may not have. Future research should investigate performance in stereotyped and non-stereotyped words to determine if this is a lexical effect.

On the other hand, /aj/-monophthongization may have been harder for participants to describe, leading to lower scores in the written task, but easier to identify in the auditory task, leading to higher scores in the auditory task, as observed for /aj/-monopthongization. This feature is highly salient, and often performed in describing a southern American accent, which suggests a high level of accuracy, as would be required for for high scores in the metalinguistic awareness task. Looking at Figure 9, /aw/-raising and /aj/-monophthongization have opposite patterns for these two tasks, which would effectively cancel out any differences that may be observed. Future research should investigate different features with varying levels of folk linguistic awareness across its various modes to determine if and how these modes contribute to different tasks for measuring metalinguistic awareness.

In conclusion, this study finds that it is possible to quantify metalinguistic awareness using

[1] This was not the case for /aj/-monophthongization since only one speaker had this feature.

[2] /aw/-raising was also the only feature for which a mean metalinguistic awareness score above 0.1 was observed in Task 1, which also supports the notion that people have a high level of accuracy for this feature.

written dialect description, written dialect identification and auditory dialect identification tasks. However, it also finds that other modes of folk linguistic awareness may contribute to results for different tasks, so these also need to be considered when quantifying metalinguistic awareness.

References

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Boersma, P. & D. Weenink. 2021. Praat: doing phonetics by computer [computer program]. version 6.1.38. http://www.praat.org/.

Chambers, Jack K. 1973. Canadian raising. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 18(2). 113–135.

Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674. https://doi.org/10.1198/106186006X133933.

Johnstone, Barbara & Scott F Kiesling. 2008. Indexicality and experience: Exploring the meanings of/aw/-monophthongization in Pittsburgh. *Journal of sociolinguistics* 12(1). 5–33.

de Leeuw, J. R. 2015. jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods* 47(1). 1–12. https://doi.org/10.3758/s13428-014-0458-y.

Onosson, Sky. 2018. *An acoustic study of canadian raising in three dialects of North American English*. Victoria, British Columbia: University of Victoria dissertation.

Plichta, Bartlomiej & Dennis R Preston. 2005. The/ay/s have it the perception of/ay/as a north-south stereotype in united states english. *Acta Linguistica Hafniensia* 37(1). 107–130.

Preston, Dennis R. 1996. Whaddayaknow?: The modes of folk linguistic awareness. *Language awareness* 5(1). 40–74.

R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. https://R-project.org/.

Ruch, Hanna. 2018. The role of acoustic distance and sociolinguistic knowledge in dialect identification. *Frontiers in psychology* 9(818). 1–15.

Stanley, Joseph A. 2018. Changes in the timber industry as a catastrophic event: BAG-raising in Cowlitz County, Washington. *University of Pennsylvania Working Papers in Linguistics* 24(2). 137–146.

Stanley, Joseph A. 2022. Regional patterns in prevelar raising. *American Speech* 97(3). 374–411.

Sullivan, Lisa. 2022. *Pre-velar /æ/-raising in Ontario and Colorado English: Production, perception and metalinguistic awareness*. Toronto, ON: Univesrity of Toronto dissertation.

Swan, Julia Thomas. 2016. Canadian English in the Pacific Northwest: A phonetic comparison of Vancouver, BC and Seattle, WA. In *Proceedings of the 2016 annual conference of the canadian linguistic association*, .

Voeten, Cesko C. 2023. *buildmer: Stepwise elimination and term reordering for mixed-effects regression*. https://CRAN.R-project.org/package=buildmer. R package version 2.9.

Zwicker, Eberhard & Ernst Terhardt. 1980. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *The Journal of the Acoustical Society of America* 68(5). 1523–1525.