INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

The Influence of Modality and Speaking Style on the Assimilation Type and Categorization Consistency of Non-Native Speech

Sarah E. Fenwick[1,2], *Catherine T. Best*[1,3], *Chris Davis*[1], *Michael D. Tyler*[1,2]

[1] MARCS Institute, Western Sydney University, Sydney, Australia
[2] School of Social Sciences and Psychology, Western Sydney University, Sydney Australia
[3] School of Humanities and Communication Arts, Western Sydney University, Sydney Australia.

s.fenwick@westernsydney.edu.au, c.best@westernsydney.edu.au,
chris.davis@westernsydney.edu.au, m.tyler@westernsydney.edu.au

Abstract

The Perceptual Assimilation Model [1] proposes that non-native contrast discrimination accuracy can be predicted by perceptual assimilation type. However, assimilation types have been based just on auditory–only (AO) citation speech. Since auditory-visual (AV) and clear speech can benefit nonnative speech perception [2, 3], we reasoned that modality and speaking style could influence assimilation. This was tested by presenting English monolinguals Sindhi consonants in a categorization task. Results showed that, across speaking styles, consonants were assimilated the same way in AV and AO. For consonants that were uncategorized in visual-only (VO) conditions: 1) their AO counterpart was more consistently categorized than AV; and 2) citation speech was also more consistently categorized than clear. Interestingly, this set of results was reversed for consonants that were assimilated to the same native category across modalities; participants were able to use the visual articulatory information to make more consistent categorization judgments for AV than AO. This was also the case for speaking style: clear speech was more consistently categorized than citation. Together these results demonstrate that the extent to which AV and clear speech is beneficial for cross-language perception may depend on the similarities between the articulatory characteristics of native and non-native consonants.

Index Terms: Sindhi consonants, cross-language perceptual assimilation, modality, speaking style.

1 Introduction

Experience with one's native language shapes the way that non-native phonemes are perceived. For the monolingual listener, a non-native phoneme is typically assimilated to the closest phonologically and/or phonetically relevant native language category. These assimilations can then be used to predict gradient levels of contrast discrimination difficulty. The addition of complementary visual articulatory information, as well as speech that is produced in a slow and exaggerated style of production to overcome a challenging communicative situation (i.e., clear speech) have been shown to be advantageous when discriminating certain cross-language contrasts [2, 3].

Perceptual assimilation research has commonly focused only on the perception of auditory-only (AO) speech that is produced for the purposes of an experimental context (i.e., citation speech) [4]. But the relationship between discrimination and assimilation may then suggest that factors which aid cross-language discrimination, i.e., auditory-visual (AV) speech [5], and clear speech [6], may also influence perceptual assimilation. According to the Perceptual Assimilation Model (PAM) [1], when a non-native phone is perceived as speech, it may be assimilated to a native category with a goodness-of-fit ranging from good to poor (*categorized*), or not categorized as any one native category (*uncategorized*). Contrast assimilation patterns are then derived from pairs of non-native phones. PAM outlines five possible ways that non-native contrasts may be assimilated: **1)** *Two-Category (TC):* Non-native phones are assimilated to two different native categories, **2)** *Single-Category (SC):* The non-native phones are assimilated to a single native category as equally good, or equally poor versions, **3)** *Category-Goodness (CG):* Both phones are assimilated to the same native category, but with varying goodness of fit, such that one phone is perceived as a better version of that native category than the other, **4)** *Uncategorised-Categorised (UC):* One non-native phone is assimilated to a native category, while the other is not, **5)** *Uncategorised-Uncategorised (UU):* Both phones fail to be assimilated to any particular native language category.

Recent research has shown that AV speech and clearly articulated speech differentially contribute to the discrimination accuracy of non-native contrasts. The discrimination of Sindhi AO and AV SC contrasts, across citation and clear speech conditions, was investigated in an AXB discrimination task with monolingual Australian English (AusE) speakers [3]. They showed an AV benefit (i.e., more accurate discrimination in AV versus AO conditions) when discriminating a non-native SC contrast that differed by place of articulation (POA) /t̪/-/ʈ/, but only in clear speech conditions. When presented with a SC contrast that differed by a laryngeal feature /b/-/ɓ/, AV benefit was found in citation but not clear speech conditions. In similar testing conditions, [2] examined the discrimination of TC and CG contrasts in speech-shaped noise. Similarly, AV benefit was found across clear and citation speech for POA TC and CG contrasts, and for laryngeal TC contrasts, but only for laryngeal CG contrasts when speech was clearly articulated. Together, these results raise the question of whether changes in discrimination may be due not only to differences in participants' ability to detect phonetic and phonological articulatory information, but also the way that these non-native speech segments are assimilated under different presentation conditions. Therefore, what remains unclear is whether the assimilation types observed from typical testing conditions (i.e., AO citation speech) remain the same across modality (i.e., AO, AV and visual-only

http://dx.doi.org/10.21437/Interspeech.2016-611

[VO]) and speaking style (i.e., clear and citation), as well as to what degree this additional articulatory information results in more systematic, or consistent, assimilation responses.

As we selected stimuli that have been shown to be categorized in AO citation speech conditions (see Section 2.2), it is predicted that there should be a similar number of categorized non-native speech segments across AO and AV conditions, but a greater number of uncategorized segments in VO conditions. Assimilation types should be similar across AO and AV conditions for TC and CG contrasts, but the addition of visual articulatory information may emphasize phonetic differences within SC contrasts, leading to some SC contrasts to shift to CG contrasts in AV and clear speech conditions. In addition, clear speech may be more consistently categorized than citation speech, and AV speech should be more consistently assimilated than both AO and VO conditions alone. By including a VO presentation condition we are able to examine the weighting, or contribution of AO and VO articulatory information when assimilating AV speech.

2 Method

2.1 Participants

Twenty-four monolingual AusE participants (18 females, 8 males, *M*age = 21.8, age range = 17-38) were recruited from first year psychology at Western Sydney University, in return for course credit.

2.2 Stimuli and Apparatus

AV speech recordings were conducted in a sound dampened booth at the MARCS Institute, Western Sydney University, by a 35-year-old female native Sindhi speaker, from Radhan, Pakistan. Multiple tokens of all 56 Sindhi consonants were recorded in phonotactically permissible /Ca/ nonsense syllables, in clear and citation speech, and were processed using the procedure described in [3]. Here we focus on 9 of those Sindhi consonants (/f, ʋ, ɗ̪, ɖ̪, ɖ, ʈ̪, ʈ, ɓ, ɓ/), which were selected from an AO citation speech categorization pre-test, where a range of categorised PAM [1] assimilation types were found [2]. In this pre-test, only citation AO tokens were presented to select stimuli based on typical PAM testing conditions. From this, two TC contrasts were selected, /f/-/ʋ/ and /ɓ/-/ɗ̪/, two CG contrasts, /ʈ̪/-/ɖ̪/ and /ɖ̪/-/ɖ/; and two SC contrasts /ʈ/-/ʈ̪/ and /b/-/ɓ/. Sindhi /ʈ̪, ʈ/ are short-lag unaspirated, and in initial position are likely to be perceived as the English *voiced* stop /d/. Sindhi /ʋ/ is a labiodental approximant likely to be perceived as the English bilabial approximant /w/.

2.3 Procedure

A categorization task was used to determine how participants assimilated foreign Sindhi consonants to their native AusE phonological categories. It was the participants' task to select one native AusE consonant category, presented visually on a grid, that best matched the Sindhi consonant they were presented (i.e., b, ch, d, f, g, h, j, k, l, m, n, p, r, s, sh, t, v, w, y, z, zh [keyword: measure], ng [keyword: hang], th [keyword: there], TH [keyword: thin]). The same speech token was then repeated and participants rated how well it matched the chosen English consonant category on a 7-point Likert scale, with '1' indicating "a very strange sounding/looking example of that category", '4' "an okay version", and '7' "a perfect example of that category". All participants completed a categorization task for each modality and speaking style, such that there were 6 tasks to complete, in randomized order: *Clear speech*: AO, AV, VO; *Citation speech*: AO, AV, VO. Within each categorization task, there were four randomized repetitions of each of the nine Sindhi consonants (/f, ʋ, ɗ̪, ɖ̪, ɖ, ʈ̪, ʈ, b, ɓ/). The duration of the testing session was 90 minutes. There were four repetitions of each token, such that each participant completed 216 trials.

3 Results

The first set of analyses assessed whether the assimilation type, as well the native language phonological category that the non-native consonants were assimilated to, was contingent on the modality and speaking style presented. Table 1 lists the L2-L1 category assimilations, categorization percentage and goodness ratings, as well as assimilation types for each speaking style, across modalities. To determine these results, the percent categorization and mean goodness-of-fit ratings (out of 7) of each Sindhi consonant category to a native language consonant category were calculated and averaged across participants. A mean percent categorization score above 50% indicated that a particular Sindhi consonant was categorized, otherwise it was deemed uncategorized [7]. When only one member of a contrast was categorized, then this resulted in a UC assimilation, but if both were uncategorized then the contrast was a UU assimilation. Where both consonant members of a contrast were categorized, then the contrast was classified as a TC, CG, or SC assimilation. To distinguish between CG and SC contrasts, where both non-native consonants were assimilated to the same native language category, an independent samples *t*-test was conducted on the mean goodness ratings for each consonant. If a significant difference was found, this was classified as a CG contrast, otherwise it was considered to be a SC assimilation. The following sections report the assimilation types for each modality in citation speech then clear speech.

3.1 Auditory-Only Citation Speech

When AO speech was presented, all non-native consonants were categorized as an AusE consonant category. Furthermore, the assimilation types found here replicate those in the stimulus selection pre-test [2]. The contrasts /f/-/ʋ/ and /ɓ/-/ɗ̪/ were TC (as English /f/-/w/ and /b/-/d/), /ʈ̪/-/ɖ̪/ and /ɖ̪/-/ɖ/ were CG, and /b/-/ɓ/ and /ʈ̪/-/ʈ/ were SC. The contrasts /ʈ̪/-/ɖ̪/ and /ɖ̪/-/ɖ/ were classified as CG assimilations because significant differences were found between the goodness-of-fit ratings of /ʈ̪/ and /ɖ̪/ to English "d", $t(31.99) = 2.68$, $p = .012$, and of /ɖ̪/ and /ɖ/ to English "d", $t(34.29) = 2.19$, $p = .036$. There were no significant differences between the goodness ratings for SC cases /b/ and /ɓ/ or /ʈ̪/ and /ʈ/.

3.2 Auditory-Visual Citation Speech

The AV L2-L1 categorization patterns, as well as the resulting assimilation types mirrored those of the citation AO conditions. The contrasts /f/-/ʋ/ and /ɓ/-/ɗ̪/ were TC assimilations; /ʈ̪/-/ɖ̪/ and /ɖ̪/-/ɖ/ as CG assimilations, and /b/-/ɓ/ and /ʈ̪/-/ʈ/ as SC assimilations. There were significant differences in the goodness ratings of /ʈ̪/ and /ɖ̪/ to English "d", $t(34.21) = 2.30$, $p = .028$, and of /ɖ̪/ and /ɖ/ to English "d", $t(37.77)$ 2.82, $p = .008$, but no significant goodness rating differences for /b/ and /ɓ/ or /ʈ̪/ and /ʈ/.

3.3 Visual-Only Citation Speech

There were only three categorized phones for VO citation speech: /f/, /ʋ/ and /b/. The contrast in which both consonants were reliably categorized across all citation modality conditions was /f/-/ʋ/, as a TC assimilation. Neither consonant within the contrasts /ɓ/-/ɗ̪/, /t̪/-/d̪/, /ʈ̪/-/ɖ̪/, or /t̪/-/ʈ̪/ were assimilated to any one particular AusE category, resulting in UU assimilations for each one. There was also one UC assimilation (/b/-/ɓ/), as /b/ was categorized to an AusE category, but /ɓ/ was uncategorized.

3.4 Auditory-Only Clear Speech

Relative to the citation AO results, three of the six assimilation types remained the same. These were the TC contrasts, /f/-/ʋ/ and /ɓ/-/ɗ̪/, the CG contrast /d̪/-/ɖ̪/ (goodness ratings: $t(46) = 2.25$, $p = .029$), and the SC contrast /b/-/ɓ/. As /t̪/ was uncategorized, the contrasts /t̪/-/d̪/ and /t̪/-/ʈ̪/ were UC assimilations.

3.5 Auditory-Visual Clear Speech

Similar to the corresponding AO versus AV citation speech responses, the clear speech AV L2-L1 categorization patterns and assimilation types were parallel to those of the clear speech AO responses. The /f/-/ʋ/ and /ɓ/-/ɗ̪/ contrasts were assimilated as TC contrasts, /d̪/-/ɖ̪/ as a CG contrast (goodness ratings: $t(46) = 2.42$, $p = .020$), and /b/-/ɓ/ as an SC contrast. The consonant /t̪/ was also uncategorized in AV conditions, and therefore /t̪/-/d̪/ and /t̪/-/ʈ̪/ were found to be UC assimilations.

3.6 Visual-Only Clear Speech

Of all clear speech conditions, VO was the condition in which the fewest Sindhi consonants were categorized. The consonants that were categorized were: /ʋ, ɓ, ɖ̪, b/. In comparison to the AO citation results, /ɓ/-/ɗ̪/ maintained a TC assimilation, and /b/ and /ɓ/ remained an SC assimilation. The consonant /f/ was uncategorized in this condition, therefore, the /f/-/ʋ/ contrast was assimilated as a UC contrast, instead of a TC contrast. All other Sindhi consonants were uncategorized, thus the contrasts /t̪/-/d̪/, /d̪/-/ɖ̪/, and /t̪/-/ʈ̪/ were classified as UU assimilations.

3.7 Categorization Consistency across Modality and Speaking Style

The second set of analyses aimed to address whether the addition of visual speech information (i.e., AV), and/or clearly articulated speech resulted in more consistent L2-L1 categorization, relative to AO citation speech alone. Therefore, both categorized and uncategorized responses were included in the analysis. Specifically, level of consistency refers to a comparison between the mean percent categorization across speaking style and modality conditions to the AusE L1 categories that were selected in typical AO citation conditions.

To assess categorization consistency, a 3 (modality: AV, AO, VO) x 2 (speaking style: clear, citation) repeated measures ANOVA was conducted. As can be seen in Figure 1, the ANOVA revealed main effects of speaking style, $F(1, 23) = 11.94$, $p = .002$, $\eta_p^2 = .34$, and modality, $F(1, 23) = 403.32$, $p < .001$, $\eta_p^2 = .95$. For speaking style, citation speech ($M = 65\%$) was more consistently categorized than clear speech ($M = 62\%$). To investigate the main effect of modality, we conducted post-hoc comparisons with a Bonferroni adjusted alpha (.05/2 = .025). This demonstrated that AO speech ($M = 80\%$) was categorized more consistently than AV ($M = 78\%$), $F(1, 23) = 6.56$, $p = .017$, $\eta_p^2 = .22$, which was more consistently categorized than VO ($M = 32\%$), $F(1, 23) = 411.95$, $p < .001$, $\eta_p^2 = .95$. The two-way interaction between speaking style and modality was not significant $F(1, 23) = .174$, $p = .68$, therefore this effect of modality did not differ significantly across the speaking styles.

The unanticipated AO consistency advantage may be due to the low number of categorized consonants in VO conditions. That is, when non-native visual articulatory information is uncategorized, and therefore not recognized as any L1 consonant category, perceivers may potentially identify the unified AO and VO (i.e., AV) segment as an incongruent percept (e.g., AO categorized as AusE 'd', while the VO counterpart is not perceived as any L1 AusE category leading

Table 1: *L2-L1 category assimilations, percent categorization (in bold), goodness ratings (in italics), and assimilation types, across speaking styles and modalities. Dark cells without values indicates that the Sindhi consonant was uncategorized. TC = Two-Category, CG = Category-Goodness, SC = Single-Category, UU = Uncategorized-Uncategorized, UC = Uncategorized-Categorized.*

	Citation Speech						*Clear Speech*					
	Auditory-Only		**Auditory-Visual**		**Visual-Only**		**Auditory-Only**		**Auditory-Visual**		**Visual-Only**	
Contrast	**L1 Category**	**Assimilation Type**	**L1 Category**	**Assimilation Type**	**L1 Category**	**Assimilation Type**	**L1 Category**	**Assimilation Type**	**L1 Category**	**Assimilation Type**	**L1 Category**	**Assimilation Type**
/f/	f **69.10%** *4.8*	TC	f **84.38%** *4.73*	TC	f **56.60%** *4.71*	TC	f **95.49%** *5.26*	TC	f **95.14%** *5.1*	TC		UC
/ʋ/	w **96.18%** *5.49*		w **99.31%** *5.66*		w **95.85%** *5.09*		w **98.61%** *5.58*		w **96.18%** *5.64*		w **73.61%** *5.08*	
/ɓ/	b **96.18%** *4.65*	TC	b **93.75%** *4.91*	TC		UU	b **90.28%** *4.72*	TC	b **99.31%** *5.7*	TC	b **75.26%** *5.47*	TC
/ɗ̪/	d **84.72%** *4.52*		d **78.47%** *4.92*				d **60.42%** *4.53*		d **50.00%** *4.77*		l **66.67%** *5.33*	
/ʈ̪/	d **60.07%** *4.6*	CG	d **54.51%** *4.77*	CG		UU		UC		UC		UU
/ɖ̪/	d **92.36%** *5.08*		d **87.15%** *5.11*				d **93.75%** *5.16*		d **90.28%** *5.24*			
/ɖ̪/	d **92.36%** *5.08*	CG	d **87.15%** *5.11*	CG		UU	d **93.75%** *5.16*	CG	d **90.28%** *5.24*	CG		UU
/ɖ/	d **71.08%** *4.5*		d **57.29%** *4.22*				d **74.65%** *4.16*		d **74.31%** *4.23*			
/b/	b **97.92%** *4.9*	SC	b **98.61%** *5.4*	SC	b **73.87%** *5.72*	UC	b **96.53%** *5.06*	SC	b **99.31%** *5.7*	SC	b **69.34%** *5.6*	SC
/ɓ/	b **96.18%** *4.65*		b **93.75%** *4.91*				b **90.28%** *4.72*		b **95.14%** *5.14*		b **75.26%** *5.47*	
/ʈ̪/	d **60.07%** *4.6*	SC	d **54.51%** *4.77*	SC		UU		UC		UC		UU
/ʈ/	d **67.13%** *4.81*		d **61.11%** *4.84*				d **52.43%** *4.94*		d **58.68%** *5.17*			

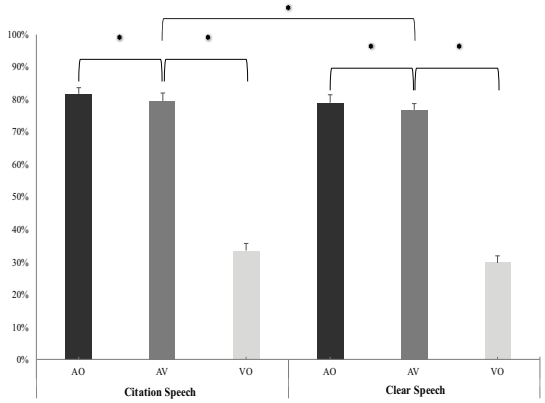

Figure 1: *Mean percent categorization of Sindhi consonants, across speaking style and modality, to English consonant categories selected in auditory-only citation conditions. AO = Auditory-Only, AV = Auditory-Visual, VO = Visual-Only. Error bars represent standard error of the mean.*

to an perceived incongruent AV percept), which has been shown to interfere with AV perception [8].

To test this prediction, a follow-up 2 (perceived match: match, mismatch) x 2 (modality: AO, AV) x 2 (speaking style: clear, citation) repeated measures ANOVA was conducted (see Figure 2). For a particular speech segment to be included within the analysis as a 'match', it was required to be categorized across modalities as the same native language category (*Citation speech:* /f, ʋ, b/; *Clear speech:* / ʋ, ɓ, ɖ̤, b/). Conversely, for a consonant to be considered a 'mismatch', it was required to be categorized to the same L1 category across AO and AV conditions, but either uncategorized or categorized to a different L1 category in VO conditions (relative to AO and AV) (*Citation speech:* /ɓ, ɖ̤, t̤, d̤, ɖ, ʈ/; *Clear speech:* /f, ɖ̤, t̤, d̤, ɖ, ʈ/).

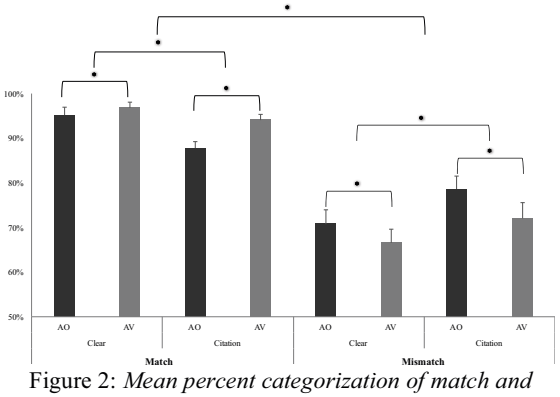

Figure 2: *Mean percent categorization of match and mismatch Sindhi consonants across each modality and speaking style. AO = Auditory-Only, AV = Auditory-Visual, VO = Visual-Only. Error bars represent standard error of the mean.*

From this analysis there was a main effect of perceived match, $F(1, 23) = 67.74$, $p < .001$, $\eta_p^2 = .75$. Predictably, matched ($M = 93\%$) speech was more consistently categorized than mismatched speech ($M = 72\%$). Modality ($F(1, 23) = 20.17$, $p < .001$, $\eta_p^2 = .47$) and speaking style ($F(1, 23) = 17.88$, $p < .001$, $\eta_p^2 = .44$) also independently interacted with matched/mismatched speech. Simple effects analyses demonstrated that for perceived mismatched consonants, AO ($M = 75\%$) was more consistently categorized than AV ($M = 69\%$), $F(1, 23) = 13.36$, $p = .001$, $\eta_p^2 = .37$. However, for perceived matched consonants, AV ($M = 95\%$) was categorized more consistently than AO ($M = 91\%$), $F(1, 23) = 18.36$, $p < .001$, $\eta_p^2 = .44$. With regard to speaking style, matched clear speech ($M = 96\%$) was more consistently categorized than citation ($M = 91\%$), $F(1, 23) = 9.80$, $p = .005$, $\eta_p^2 = .30$, but for mismatched, citation speech ($M = 75\%$), was more consistently categorized than clear ($M = 69\%$), $F(1, 23) = 11.22$, $p = .003$, $\eta_p^2 = .33$. The three-way interaction was not significant, $F(1, 23) = 2.47$, $p = .13$.

4 Discussion

The purpose of this study was to test whether assimilation type and categorization consistency is influenced by modality and speaking style. In terms of the L2-L1 assimilations, and resulting assimilation types, there were four main observations: **1)** For both speaking styles, assimilation types were the same across AO and AV conditions. However, clear speech resulted in /t̤/ becoming uncategorized; **2)** L2-L1 assimilation were poorest in VO conditions, presumably due to there being fewer distinguishable visemes than phonemes [9]; **3)** TC phonological judgments were more robust than other assimilation types across modality and speaking style; **4)** When an L2 consonant was categorized, it was to the same native language AusE phonological category across AO, AV and VO. There was one exception to this finding, where in AO and AV conditions /ɖ̤/ had been categorized as AusE /d/, but in clear speech VO conditions it was assimilated to AusE /l/. This is not surprising as both /d/ and /l/ share the same place of articulation (coronal, i.e., tongue tip contact near alveolar ridge).

The assimilation consistency results provide important information concerning the articulatory information that monolingual perceivers are able to take advantage of. Specifically, when VO speech is assimilated to the same L1 category as its AO and AV counterparts (matched), perceivers make more consistent categorization judgments in AV than AO. But, when VO information is ambiguous, or assimilated to a different L1 category (mismatched), then poorer AV than AO performance is found. These differences found between perceived matched versus mismatched assimilation also modulates the effect of speaking style. Clear speech is only beneficial to assimilation consistency when the L2 consonant is categorized as the same L1 category across all modalities (matched). If clear-spe ech articulation made VO information ambiguous, then participants were more consistent when categorizing citation speech.

A possible explanation for these results is the degree of 'native-likeness' between the AO and VO articulatory inventories of the L1 and L2. For both AO and VO speech, when an L2 phoneme is perceived as an L1 exemplar, and therefore perceived as *native-like*, perceivers are able to use this articulatory information to their advantage when assimilating AV speech. But, when there is no perceived L1 counterpart, and the L2 consonant is therefore perceived as *less native-like*, or when the L2 is perceived as a different L1 counterpart, AV consistency is hindered. Similarly, it is only when the exaggerated articulatory gestures of clear speech (across modalities) mirrors an L1 category that clear speech is beneficial (see also [10]). Therefore, future research will seek to extend these findings by examining the perceptual assimilation of AO, AV and VO clear and citation speech, for other target language and L1 participant groups, based on their articulatory phonemic and visemic inventory similarities and differences. This may also provide insights relevant to L2 learner instruction, as teaching strategies may need to be customized dependent on L2-L1 patterns.

5 References

[1] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech perception and linguistic experience: Issues in Cross-Language Research*, W. Strange, Ed. Baltimore: York Press, 1995, pp. 171–206.

[2] S. E. Fenwick, C. Davis, C. T. Best, and M. D. Tyler, "The Effect of Modality and Speaking Style on the Discrimination of Non- Native Phonological and Phonetic Contrasts in Noise," in *Proceedings of the 1st Joint Conference on Facial Analysis, Animation, and Auditory-Visual Speech Processing*, 2015.

[3] S. E. Fenwick, C. T. Best, and M. D. Tyler, "Non-Native Discrimination Across Speaking Style, Modality and Phonetic Feature," in *Proceedings of the 18th International Congress of Phonetic Sciences*, 2015.

[4] T. L. Face, "Intonation in Spanish declaratives: differences between lab speech and spontaneous speech," *Catalan J. Linguist.*, vol. 2, pp. 115–131, 2003.

[5] M. L. G. Lecumberri, M. Cooke, and A. Cutler, "Non-native speech perception in adverse conditions: A review," *Speech Commun.*, vol. 52, no. 11–12, pp. 864–886, Nov. 2010.

[6] J. Gagné, A. Rochette, and M. Charest, "Auditory, visual and audiovisual clear speech," *Speech Commun.*, vol. 37, pp. 213–230, 2002.

[7] M. D. Tyler, C. T. Best, A. Faber, and A. G. Levitt, "Perceptual assimilation and discrimination of non-native vowel contrasts.," *Phonetica*, vol. 71, no. 1, pp. 4–21, Jan. 2014.

[8] T. Paris, J. Kim, and C. Davis, "Visual Speech Speeds Up Auditory Identification Responses," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[9] B. E. Walden, R. A. Prosek, A. A. Montgomery, C. K. Scherr, and C. J. Jones, "Effects of training on the visual recognition of consonants," *J. Speech, Lang. Hear. Res.*, vol. 20, no. 1, pp. 130–145, 1977.

[10] A. R. Bradlow and T. Bent, "The clear speech effect for non-native listeners," *J. Acoust. Soc. Am.*, vol. 112, no. 1, p. 272, 2002.