# Individual variability in cue weighting for first-language vowels 

Payam Ghaffarvand Mokari and Stefan Werner<br>University of Eastern Finland<br>payam.ghaffarvand@uef.fi ORCID: https://orcid.org/0000-0002-1816-2783<br>stefan.werner@uef.fi ORCID: https://orcid.org/0000-0001-5176-8114

Submitted: 21/01/2017. Accepted: 07/07/2017. Published online: 21/02/2018
Citation / Cómo citar este artículo: Ghaffarvand Mokari, P. and Werner, S. (2017). Individual variability in cue weighting for first-language vowels. Loquens, 4(2), e044. doi: http://dx.doi.org/10.3989/loquens.2017.044


#### Abstract

This study investigates the use of different cues in discrimination of Azerbaijani $/ \rightsquigarrow / \mathrm{and} / \mathrm{m} / \mathrm{vowels}$. Regarding the large overlap in $f_{1}-f_{2}$ vowel space in the production of these vowels, this study researched for other possible cues in their categorization. Twenty native Azerbaijani listeners were tested through a perceptual identification test. Since $f_{2}$ was weighted more consistently through the experiment, it is suggested that $f_{2}$ is the primary cue in discriminating of this vowel pair. We observed individual differences in the perceptual weighting of $f_{2}$ and $f_{3}$ among the listeners. Although most of the participants gave more weight to $f_{2}$, some others weighted $f_{3}$ heavier than $f_{2}$ or gave weight to both cues equally. These findings expand the knowledge on perceptual cue weighting and point the importance of examining cue weighting at the individual level.


Keywords: individual differences; cue weighting; perception; Azerbaijani vowels.
RESUMEN: Variabilidad individual en el peso de claves en vocales de una primera lengua.-Este estudio investiga el uso de diferentes claves en la discriminación de las vocales $/ œ / \mathrm{y} / \mathrm{u} /$ del azerbaiyano. Teniendo en cuenta el considerable solapamiento que en la producción de estas vocales se produce en el espacio vocálico de $f_{1}-f_{2}$, en este estudio se han investigado otras posibles claves en su categorización. Veinte oyentes, hablantes nativos de azerbaiyano, participaron en un test perceptivo de identificación. Puesto que se otorgó a $f_{2}$ un peso más consistente a lo largo del experimento, se sugiere que $f_{2}$ es la clave primaria para identificar este par vocálico, si bien observamos diferencias individuales entre los hablantes en el peso atribuido a $f_{2}$ y $f_{3}$. Aunque la mayor parte de los participantes otorgaron más peso a $f_{2}$, otros se lo atribuyeron a $f_{3}$ en mayor medida que a $f_{2}$, o se apoyaron en las dos claves por igual. Estos resultados amplían el conocimiento que se posee sobre el peso de las distintas clases perceptivas y ponen de manifiesto la importancia que tiene examinar el peso de tales claves en cada individuo.

Palabras clave: diferencias individuales; peso de las claves; percepción; vocales del azerbaiyano.

## 1. INTRODUCTION

There are multiple acoustic dimensions that define speech categories. During spoken language comprehension, listeners categorize speech sounds based on these continuous acoustic cues. Listeners need to determine which cues are relevant to pay attention to, and what relative importance each cue has in order to assign more weight to that cue. Several studies have attempted to find the acoustic dimensions that are important in the discrimination of different speech sounds. Morrison (2013) provides a review of theories related to dynamic aspects of vowel perception. Strange and Jenkins (2013), in their Dynamic Specification model, mention that the most im-
portant cues to vowel identity are in the spectro-temporal patterns of consonant-vowel and vowel-consonant formant transitions.

Among the early studies on the role of different cues in the perception of vowels is the study by Bennett (1968), which investigated the relative importance of the spectral and temporal cues in the discrimination of pairs of English and German vowels. He suggested that the importance of the temporal cue is inversely proportional to the distance between the qualities of a given pair of vowels. His results showed that spectral form is, in general, more important than duration in vowel recognition in both English and German, and it is only when two vowels are very close in quality that the duration cue is more important for their discrim-
ination. Ainsworth (1972) used sets of synthetic vowel sounds that differed in first-formant frequency, second-formant frequency, and duration, and investigated the effect of these cues on the identification of vowels. He found that listeners' judgments depended on all of these factors; however, duration was a relatively more important cue for vowels located in the centre of the $f_{1}-f_{2}$ space where a vowel might more readily be confused with one of its neighbours.

According to Idemaru et al. (2012), "whereas any of the acoustic dimensions may play a role in phonetic categorization, they are not necessarily perceptually equivalent". Giving greater perceptual weight to some of the acoustic dimensions is referred to as cue weighting (Holt \& Lotto, 2006; Francis, Kaganovich, \& Driscoll-Huber, 2008; Idemaru et al., 2012). Hillenbrand, Clark, and Houde (2000) found that English listeners give more weight to the spectral than the temporal dimension in categorizing English [i] and [I] vowels. It has also been found that in discrimination of voiced and voiceless bilabial stop consonants at the syllable initial position, voice onset time (VOT) is more strongly weighted by English listeners and fundamental frequency $\left(f_{0}\right)$ of the following vowel is used as a secondary cue in the discrimination of this pair (Abramson \& Lisker, 1985; Francis et al., 2008). Holt and Lotto (2006) suggest that dimensions that are highly related to category identity need to be more strongly perceptually weighted than those less predictive of category identity. These acoustic dimensions are sometimes weighted differently among listeners.

There is not extensive research on individual differences in cue-weighting strategies in speech perception (see, e.g., Allen, Miller, \& DeSteno, 2003; Haggard, Ambler, \& Callow, 1970; Hazan \& Rosen, 1991; Idemaru, Holt, \& Seltman, 2012; Kong \& Edwards, 2011, 2016; Raizada, Tsao, Liu, \& Kuhl, 2010; Shultz, Francis, \& Llanos, 2012). Some studies have shown that individual variability exists in cue-weighting strategies (Haggard et al., 1970; Hazan \& Rosen, 1991; Idemaru et al., 2012). For instance, Shultz et al. (2012) reported individual differences in the extent of reliance on the secondary cue $\left(f_{0}\right)$ in discrimination of $/ \mathrm{b} /$ and $/ \mathrm{p} /$. Idemaru et al. (2012) suggest that examining individual differences in perceptual cue weighting in a situation where different dimensions provide similar informativeness provides an opportunity to better understand listeners' sensitivity to distributional characteristics of acoustic dimensions in categorization of speech sounds. For instance, Chládková, Hamann, Williams, and Hellmuth (2016) found that $f_{2}$ slope direction is used as a cue (additional to midpoint formant values) to distinguish /i:/ from /u:/ by British English listeners.

### 1.1. Vowel perception

Through the history of speech research, formants have played an important role in the studies of vowel perception and acoustic descriptions of vowels. Fant (1960) indicated the importance of formant frequencies as the prime determinants of the spectral envelope of oral vowels, suggesting that the complex spectra of vowel-like
sounds could be uniquely indexed with relatively few parameters. Since formant amplitude appeared to be redundant with formant frequency (Fant, 1956; Stevens, 1998) and because formant bandwidth appeared to have little influence on perception (Klatt, 1982), the focus in speech perception studies was placed on formant frequencies as correlates of perceptual vowel identification.

Some other studies have shown that general spectral shape is a good correlate to the measures of psychoacoustic distance between vowel-like stimuli (Bladon \& Lindblom, 1981; Pols, van der Kamp, \& Plomp, 1969). In this approach, it is suggested that listeners compare vowel spectra to find the closest match to an internal representation of the corresponding vowel categories. Therefore, formant peaks are not treated differently from other spectral properties and all spectral components are given weight. However, Kiefte and Kluender (2008) proposed that listeners ignore spectral shape properties in the identification of synthetic monophthongs when the target stimuli were embedded in a sentence.

In addition to formant-based and spectral shape approaches another approach in vowel perception research is the concept of spectral features based on an intermediate representation. Some studies mention that auditory $f_{2}$ and $f_{3}$ are perceptually interrelated. Delattre, Liberman, Cooper, and Gerstman (1952) noted that it was possible to produce acceptable versions of French vowels with the Pattern Playback, with two energy bands close to measured $f_{1}$ and $f_{2}$ of naturally produced vowels. This higher $f_{2}$ has been called the effective $\mathrm{f}_{2}$ or $\mathrm{f}_{2}$ prime (Fant, 1973; Fant \& Risberg, 1963). Chistovich and Lublinskaya (1979) proposed that formant peaks closer than 3.0-3.5 Bark are merged into a single perceived spectral prominence.

Fujimura (1967) studied the perception of high vowels in Swedish to investigate the theories of formant integration into $f_{2}$ prime. The notion of wideband integration of $f_{2}$ and $f_{3}$ was criticized based on his results and he instead proposed that both $f_{2}$ and $f_{3}$ make independent contributions even when they are separated by less than 3.0 Bark. Rosner and Pickering (1994, pp. 151-152) give experimental evidence indicating that it is not likely that higher formants merge auditorily into a single effective perceptual feature. Neary and Kiefte (2003) used a neural network to model spectral integration similar to that proposed by $f_{2}$ prime models in order to reduce a large threedimensional vowel continuum to two effective formants or parameters; however, this attempt was not successful. A three-dimensional formant-based representation performed substantially better in predicting listeners' vowel judgments than any two-dimensional representation that could be discovered with the neural network. This again supports Fujimura's (1967) hypothesis that vowel perception cannot be explained with two parameters alone.

### 1.2. Azerbaijani vowels

The Azerbaijani belongs to the western group of the southwestern, or Oghuz, branch of the Turkic language
family and is mainly spoken in Azerbaijan and Iran. Among nonPersian languages in Iran, Azerbaijani, with approximately $15-20$ million native speakers, has the largest number of speakers (Crystal, 2010). Azerbaijani has nine vowels, /æ a o e œ u u i y/, with no length distinction (Figure 1).

### 1.3. Present study

Most of the vowels are acoustically differentiated in terms of their first and second formant $\left(f_{1}\right.$ and $\left.f_{2}\right)$ values in different languages. In a recent research, Ghaffarvand Mokari and Werner (2016) found a large overlap in $f_{1}-f_{2}$ space between Azerbaijani / $\mathrm{m} /$ and /œ/ vowels (Figures 2 and 3). The $/ \mathrm{m} /$ and $/ œ /$ vowel are contrastive phonemes in different contexts in Azerbaijani (e.g., /suz/ 'groan' versus /sœz/ 'word').

Figure 1: A vowel chart of Azerbaijani (Ghaffarvand Mokari \& Werner, 2017).


Figure 2: Distribution of the Azerbaijani vowels in $f_{1} \times f_{2}$ (Bark) based on production of the 23 female participants in the study by Ghaffarvand Mokari and Werner (2016). The ellipses representing two standard deviations from the mean.


Linear discriminant analysis revealed that $f_{1}$ and $f_{2}$ as predictors fail to accurately classify these two vowels. Further inclusion of $f_{0}$ and duration as the predictors also did not improve the classification percentage. However, the inclusion of $f_{3}$ to the predictors improved the classifications pretty dramatically. It seems these two vowels are more distinct based on $f_{3}$ information (Figure 4). Holt and Lotto (2006) argue that "if there is not much overlap in a specific acoustic dimension, then that dimension would be very informative about category identity and it would be expected to receive more perceptual weight than the other acoustic dimensions".

Figure 3: Scatterplot of /uu/ and /œ/ vowels based on production of the 23 female participants in the study by Ghaffarvand Mokari and Werner (2016).

Axes represent $f_{1}, f_{2}$ values in Hz .


Figure 4: 3D scatterplot of /u/ and /œ/ vowels based on the productions of 23 female participants in the study by Ghaffarvand Mokari and Werner (2016). Axes represent $f_{1}, f_{2}$, and $f_{3}$ values in Hz .


Based on the hypothesis of Fujimura (1967) and findings of Chistovich and Lublinskaya (1979), since the difference between $f_{2}$ and $f_{3}$ at least in Azerbaijani / $\mathbf{w} /$ vowel is more than 3.5 Bark (Ghaffarvand Mokari \& Werner, 2016), in this study we assumed $f_{2}$ and $f_{3}$ as separate perceptual parameters. Studies on how listeners weight perceptual cues in categorization of L1 vowels are limited, especially when vowels are differentiated only by spectral features. We designed the present study in order to explore the perceptual categorization of the two Azerbaijani $/ \mathrm{m} /$ and $/ \rightsquigarrow /$ vowels. We first aim to find out how listeners weight $f_{2}$ and $f_{3}$ in discriminating the $/ \mathrm{m} /-/ \propto /$ pair. We aim to find whether $f_{3}$ is an important perceptual cue in the discrimination of this pair or not. We are specifically interested in observing how listeners weigh different acoustic dimensions of these vowels and if they use different cue-weighting strategies in their discrimination.

## 2. METHOD

### 2.1. Participants

Participants were 10 male and 10 female native Azerbaijani speakers in Tabriz, north-west of Iran. They were born and grown in Tabriz, always used Azerbaijani as the communication language, and reported no history of hearing or other speech problems. They had a mean (SD) age of 30.4 (5.4) years. An informed consent form was obtained from all participants.

### 2.2. Stimuli

A 29-year-old male native speaker of Azerbaijani from Tabriz produced several examples of the word /bœl/ 'divide' in isolation. All tokens were recorded in a soundtreated room using a ZOOM H6 recorder positioned at approximately 20 cm in front of the speaker. Recordings were at a sampling rate of 44.1 kHz with 16bit resolution. One natural production of the token /bœl/ was selected. The selected token had no sudden changes in formants during the periodic portion of the signal, no changes in fundamental frequency, and no clicks. For the resynthesis of the tokens, a periodic portion of the vowel wave form was manually extracted from the /bœl/ token, from the end of the $/ \mathrm{b} /$ burst to the last zero crossing of the vowel waveform before the silent gap. The first three formants $\left(f_{1}, f_{2}\right.$, and $\left.f_{3}\right)$ were measured using the standard LPC analysis in Praat (version 6.0.21). The first formant was 475 Hz , the second formant was 1386 Hz , and the third formant was 2273 Hz . The average intensity was 70 dB . We synthesized this token in Praat and created 24 stimuli for the perception experiment. Three sets of tokens were made: (1) by only manipulating the $f_{3}$, (2) by only manipulating the $f_{2}$, (3) by manipulating $f_{2}$ and $f_{3}$. For each set, eight spectral steps (equal along a bark scale [1 step $=0.22$ Bark for $f_{2}$ and 1 step $=0.19$ Bark] for $f_{3}$ ) were created (Figure 5).

Figure 5: Stimuli $f_{2}$ and $f_{3}$ values.


We decided to use extreme spectral values within the one standard deviation of the mean for these vowels as absolute exemplars. The values of the two stimuli are based on mean values of the Azerbaijani vowels /œ/ and $/ \mathrm{u} /$ in productions of male speakers reported by Ghaffarvand Mokari and Werner (2016). The absolute /œ/like instance was the token with the highest $f_{2}$ and the lowest $f_{3}$ (the upper-right corner in Figure 5) and the absolute /um/ like instance was the token with the lowest $f_{2}$ and the highest $f_{3}$ in the continuum (the lower-left corner in Figure 5). Additionally, we asked four Azerbaijani native listeners to approve if synthetic tokens were the exemplars of the intended vowels to use in the experiment.

On each trial of the test in a XAB task, three vowel tokens were played and the listeners were asked to decide whether the first vowel sounded like the second (A) or third (B). The second and third tokens were the most /œ/ like and $/ \mathrm{m} /$ like stimulus, and the first token was one of the 24 stimuli. Participants had to classify each of the 24 stimuli as one of the two absolute exemplars of the vowels. Following Werker and Logan (1985) and Escudero, Benders, and Lipski (2009) the interval between the three tokens was set to 1.2 seconds in order to ensure lan-guage-specific phonological processing. The order of the presentation of the A and B stimuli was counterbalanced, leading to 48 different XAB trials, which were presented four times each. This way, we ended up with a total of 192 trials.

### 2.3. Procedure

The listeners were tested individually in a quiet room by native Azerbaijani speakers who gave all instructions in Azerbaijani. Prior to the experiment, the absolute exemplars of the vowels (the most/œ/like token and the most $/ \mathrm{m} /$ /like token of the stimulus set) were played and participants were asked to pronounce each endpoint and
mention three words that include that vowel. This was to ensure that the tokens were easily identifiable as the intended Azerbaijani vowels by native Azerbaijani listeners (Escudero et al., 2009). The listeners were asked to click on a computer screen displaying the numbers " 1 ", " 2 ", and " 3 ". The number " 1 " was presented in grey color and was non-clickable. The test was carried out on a PC using Praat. The experiment lasted approximately 20 minutes for each participant. They had a 5minute break in the middle of the experiment. All participants' accuracy in discrimination of the absolute exemplars of either Azerbaijani $/ œ /$ or $/ \mathrm{m} /$ was more than $80 \%$, so we did not exclude any of them.

### 2.4. Analysis

We performed logistic regression analysis to investigate the listeners' use of different spectral information. The equation (1) is the example equation used for a model including $f_{2}$ and $f_{3}$.

$$
\begin{gather*}
\log (\mathrm{odds}(œ))=\ln (p(œ) / p(\mathrm{u}))=  \tag{1}\\
=\alpha+\beta \mathrm{f} 2 \times f_{2} \text { values }+\beta \mathrm{f} 3 \times f_{3} \text { values }
\end{gather*}
$$

In this equation, $\alpha$ is the intercept of the regression model. The coefficients ( $\beta$ 's) show how much a one-step difference in one of the predictors makes a change in the $\log$ odds of a participant's response. Hence, according to the suggestions by Morrison (2007, 2009), $\beta$ is regarded as participant's reliance on each of these cues. Following Escudero et al. (2009) we used equation (2) to compute the relative reliance of the participants on each cue. Values higher than 0.5 mean that $f_{2}$ is weighted heavier than $f_{3}$ and those below 0.5 show that $f_{3}$ is weighted heavier.

$$
\begin{equation*}
\text { cue weighting }=\beta_{f 2} /\left(\beta_{f 2}+\beta_{f 3}\right) \tag{2}
\end{equation*}
$$

Also, as mentioned by Escudero et al. (2009), polarcoordinate magnitude can be calculated using the logistic regression coefficients, which indicate the boundary crispness. The larger polar-coordinate magnitude indicates a clearer boundary between the two categories (Morrison, 2007). The polar-coordinate magnitude for the model including only $f_{2}$ was measured as indicated in equation (3).
(3) polar-coordinate magnitude $=\sqrt{\beta_{\mathrm{f} 2}{ }^{2}}$

Based on the results of a logistic regression analysis, it is also possible to find whether an individual cue significantly affects a listener's responses or not. To this end, we tested whether a logistic regression model that includes a cue as an independent variable predicts the responses significantly better than the null model. For instance, the effect of $f_{2}$ is evaluated through the comparison of the fit of a model with $f_{2}$ as a predictor to the fit of a model with only the intercept. The fit difference between the two models is the $\Delta G^{2}$, which is approximately $\chi^{2}$ dis-
tributed. The difference in degrees of freedom between the two models are the degrees of freedom of this $\Delta G^{2}$. These results are reported using an $\alpha$ level of 0.05 for each participant.

## 3. RESULTS

A series of $\mathrm{G}^{2}$ comparisons as described in the method section were performed to examine which of the cues were used significantly for vowel categorization. Inclusion of $f_{2}$ significantly improved the fit of the model for 20 out of 20 participants ( $p<0.05$ ), compared to a model without any independent variable; when only $f_{3}$ was included, the fit of the model significantly improved for 9 participants out of $20(p<0.05)$, compared to a model without any factor. Figure 6 represents a scatterplot for the coefficients of the regression model with $f_{2}$ and $f_{3}$ for the 20 participants. Escudero et al. (2009) mention that "the coefficients of the logistic regression analysis show to what extent a one-step difference in one of the predictors causes a change in the log odds of a participant's response (p. 457)".

As described in 2.4 , a cue weighting of 0.5 indicates that the listener weights both cues equally, and a cue weighting higher than 0.5 indicates that the weighting of $f_{2}$ is heavier than that of $f_{3}$; if it is below $0.5, f_{3}$ is weighted heavier. Figure 7 represents the mean cue weighting for each of the participants. The closer the coefficients of $f_{2}$ and $f_{3}$ (Figure 6) are to each other, the closer the relative cue weighting is to the center line (both cues weighted equally; Figure 7).

According to Figure 7, most of the listeners weighted $f_{2}$ heavier, and some listeners weighted $f_{3}$ heavier or both equally. However, there are differences on the amount of weighting of cues among the listeners. Overall, the reliance on $f_{2}$ was much stronger than on $f_{3}$. Some partici-

Figure 6: Scatterplot of coefficients from the logistic regression analysis that shows the reliance on $f_{2}$ and $f_{3}$.


Figure 7: Relative cue weighting of $f_{2}$ and $f_{3}$ per participant.

pants' relative cue-weighting scores were close to $1\left(f_{2}\right.$ only); however, the three participants who gave more weight to $f_{3}$ did not weight it as strong as the average of participants who gave more weight to $f_{2}$.

Finally, we computed the polar-coordinate magnitudes of the two models (only $f_{2}$ and $f_{2}+f_{3}$ ) and compared the models' steepness in categorization boundaries. As mentioned by Morrison (2007),
the contrast coefficient slope in the logistic space is related to the slope of the sigmoidal curve which represents the rate of change from one category to another in the probability space. The size of the contrast coefficient and the corresponding steepness of the steepest tangent to the sigmoidal curve in the probability space are indicators of the crispness of the boundary between the two categories" (p.229).

Figure 8 shows the probability of choosing the /œ/ vowel along the 8 steps. Compared to the model when only $f_{2}$ changes toward the / $</$ vowel, changes of both in $f_{2}$ and $f_{3}$ make the probability of $/ \rightsquigarrow /$ response to jump steeper toward 1.

There was a significance difference between the coefficients of these two models ( $t=-2.03, p=0.05$ ), and inclusion of $f_{3}$ made the curve steeper compared to model with only $f_{2}$ (Figure 8). The mean polar-magnitude values for models were $f_{2}+f_{3}=0.83>f_{2}=0.74$.

## 4. DISCUSSION

The current study examined perceptual weighting of different acoustic dimensions in perceptual discrimination of Azerbaijani $/ \rightsquigarrow /$ and $/ \mathrm{u} /$ / vowels. Regarding the large overlap in $f_{1}-f_{2}$ vowel space in the production of these two vowels (Ghaffarvand Mokari \& Werner, 2016), this study explored if other cues play a role in their discrimination. To the best of our knowledge, this is the first study of cue weight in the discrimination of native vowels

Figure 8: Sigmoidal curves in the probability space for contrast coefficient values of model $f_{2}$ and $f_{2}+f_{3}$.

based solely on spectral information. Our results revealed individual differences in the cue weighting for the perception of Azerbaijani /œ/ and /ui/ vowels. Although $f_{2}$ was a more important cue for the discrimination of this vowel pair, $f_{3}$ also played a role. Our results of the reliance on different cues revealed that a higher number of listeners mostly relied on $f_{2}$ and some of them relied on $f_{3}$ or on both.

Overall, one of the important findings in the present study is that $f_{2}$ is still the main cue in the distinction of these two vowels despite their large overlap regarding their $f_{2}$ values. One explanation for this issue would be that listeners use a perceptual vowel-intrinsic normalization process that does not need the information from other vowels. According to Adank et al. (2004) "vowel-intrinsic normalization models have been considered to be
more suitable as models for human vowel perception" because they "can normalize a single vowel from a speaker without information about other vowels from that speaker" (p. 3105).

These findings are in line with individual differences reported in previous studies. Regarding the discrimination of voiced and voiceless stops, Stevens and Klatt (1974) report that some listeners relied more on VOT than $f_{1}$ onset frequency, and Haggard et al. (1970) found that some listeners were more sensitive to $f_{0}$ than VOT in distinguishing voiced and voiceless stops. More recently, Idemaru et al. (2012) found considerable variability between Japanese listeners' perceptual weighting of absolute and relative durations in the discrimination of Japanese singleton and geminate stop categories.

In addition, our results revealed that the categorical boundary was steeper when both $f_{2}$ and $f_{3}$ were included in the model compared to the model including only $f_{2}$. This was in line with the study by Hazan and Rosen (1991), who observed that listeners' identification functions were uniformly steep in the full-cue condition.

Some previous studies have indicated that $f_{2}$ and $f_{3}$ might be perceptually regarded as one percept $\left(f_{2}^{2}\right.$ prime; Delattre et al., 1952; Fant \& Risberg, 1963; Fox, Jacewicz, \& Chang, 2011). Chistovich and Lublinskaya (1979) proposed that close formant peaks are merged into a single perceived spectral prominence. However, other studies suggest that a two-dimensional view on formants cannot explain vowel perception and at least three spectrally prominent regains (corresponding to $f_{1}, f_{2}$, and $f_{3}$ ) are necessary to explain vowel perception (Fujimura, 1967; Rosner \& Pickering, 1994). There is a need for further studies to investigate whether $f_{2}$ and $f_{3}$ covary in perceptual discrimination of Azerbaijani $/ œ /$ and $/ \mathrm{m} /$ vowels and whether they can be merged into one dimension in perception or not.

Francis et al. (2008) suggest that listeners normally rely on primary cues (e.g., on VOT in the discrimination of English stop voicing contrast) in ideal listening conditions. However, they adjust their cue weighting toward secondary cues under less-than-ideal conditions, for instance when listening to speech in noise or listening to multiple speakers. If $f_{2}$ is the primary cue in the discrimination of the Azerbaijani /œ/ and /u/ vowels, it can be assumed that listeners will use $f_{3}$ cue in noisy and not in ideal conditions.

Our results also revealed individual differences in categorization gradiency in presence of different cues. In explanation of individual differences in speech perception, Kong and Edwards (2016) hypothesized that gradiency would be related to general cognitive control. They tested this hypothesis by correlating measures of gradiency with performance on measures of inhibition and task shifting and found little support for this claim. Kapnoula (2016) also did not find consistent relationships between gradiency and measures of executive function. Idemaru et al. (2012) speculate that their observed individual cueweighting pattern can be due to the similar informativeness of the acoustic dimensions and it allows listeners to
freely use either source of information, perhaps varying in which information they use across time.

Future research may look into the relation between production and perception of reliance on spectral cues. One would assume that the individuals who give more weight to $f_{3}$ in discrimination of the Azerbaijani/œ/ and $/ \mathrm{um} /$ vowels produce them also with heavier $f_{3}$ differences. In summary, we observed some individual differences in cue-weighting strategies among native listeners. Although there are a few studies on the individual differences in cue weighting, the source of these differences still remains to be discovered in future.

## REFERENCES

Abramson, A. S., \& Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. Fromkin (Ed.), Phonetic linguistics: Essays in honor of Peter Ladefoged (pp. 25-33). New York: Academic.
Adank, P., Smits, R., \& Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. The Journal of the Acoustical Society of America, 116(5), 3099-3107. https://doi.org/10.1121/1.1795335
Ainsworth, W. A. (1972). Duration as a cue in the recognition of synthetic vowels. The Journal of the Acoustical Society of America, 51, 648-651. https://doi.org/10.1121/1.1912889
Allen, J. S., Miller, J. L., \& DeSteno, D. (2003). Individual talker differences in voice-onset-time. The Journal of the Acoustical Society of America, 113, 544-552. https://doi.org/10.1121/1.1528172
Bennett, D. (1968). Spectral form and duration as cues in the recognition of English and German vowels. Language and Speech, 11(2), 65-85. https://doi.org/10.1177/002383096801100201
Bladon, R. A. W., \& Lindblom, B. (1981). Modeling the judgment of vowel quality differences. The Journal of the Acoustical Society of America, 69, 1414-1422. https://doi.org/10.1121/1.385824
Chistovich, L. A., \& Lublinskaya, V. V. (1979). The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. Hearing Research, 1(3), 185-195. https://doi. org/10.1016/0378-5955(79)90012-1
Chládková, K., Hamann, S., Williams, D., \& Hellmuth, S. (2016). F2 slope as a perceptual cue for the front-back contrast in Standard Southern British English. Language and Speech, 60(3), 377-398. https://doi.org/10.1177/0023830916650991
Crystal, D. (2010). The Cambridge encyclopedia of language (3rd ed.). Cambridge and New York: Cambridge University Press.
Delattre, P., Liberman, A. M., Cooper, F. S., \& Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color; Observations on one-and two-formant vowels synthesized from spectrographic patterns. Word, 8(3), 195-210. https://doi.org/10.1080/00437956.1952.11659431
Escudero, P., Benders, T., \& Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. Journal of Phonetics, 37(4), 452-465. https://doi.org/10.1016/j.wocn.2009.07.006
Fant, G. (1956). On the predictability of formant levels and spectrum envelopes from formant frequencies. In M. Halle (Ed.), For Roman Jakobson (pp. 109-120). The Hague: Mouton.
Fant, G. (1960). Acoustic theory of speech production. Mouton: The Hague.
Fant, G. (1973). Speech sounds and features. Cambridge: MIT Press.
Fant, G., \& Risberg, A. (1963). Auditory matching of vowels with two formant synthetic sounds. STL-Quarterly Progress Status Report, 4(4), 7-11.
Fox, R. A., Jacewicz, E., \& Chang, CY. (2011). Auditory spectral integration in the perception of static vowels. Journal of Speech, Language, and Hearing Research, 54, 1667-1681. https://doi.org/10.1044/1092-4388(2011/09-0279)

Francis, A. L., Kaganovich, N., \& Driscoll-Huber, C. (2008). Cuespecific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. The Journal of the Acoustical Society of America, 124, 1234-1251. https://doi.org/10.1121/1.2945161
Fujimura, O. (1967). On the second spectral peak of front vowels: A perceptual study of the role of the second and third formants. Language and Speech, 10(83), 181-193. https://doi.org/ 10.1177/002383096701000304

Ghaffarvand Mokari, P., \& Werner, S. (2016). An acoustic description of spectral and temporal characteristics of Azerbaijani vowels. Poznań Studies in Contemporary Linguistics, 52(3), 503-518. https://doi.org/10.1515/psicl-2016-0019
Ghaffarvand Mokari, P., \& Werner, S. (2017). Azerbaijani. Journal of the International Phonetic Association, 47(2), 207-212. https://doi.org/10.1017/S0025100317000184
Haggard, M., Ambler, S., \& Callow, M. (1970). Pitch as a voicing cue. The Journal of the Acoustical Society of America, 47, 613617. https://doi.org/10.1121/1.1911936

Hazan, V., \& Rosen, S. (1991). Individual variability in the perception of cues to place contrasts in initial stops. Attention, Perception, \& Psychophysics, 49(2), 187-200. https://doi.org/10.3758/ Bf03205038
Hillenbrand, J. M., Clark, M. J., \& Houde, R. A. (2000). Some effects of duration on vowel recognition. The Journal of the Acoustical Society of America, 108, 3013-3022. https://doi. org/10.1121/1.1323463
Holt, L. L., \& Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. The Journal of the Acoustical Society of America, 119, 30593071. https://doi.org/10.1121/1.2188377

Idemaru, K., Holt, L. L., \& Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. The Journal of the Acoustical Society of America, 132, 3950-3964. https://doi.org/10.1121/1.4765076
Kapnoula, E. E. (2016). Individual differences in speech perception: sources, functions, and consequences of phoneme categorization gradiency ( PhD thesis). University of Iowa.
Kiefte, M., \& Kluender, K. R. (2008). Absorption of reliable spectral characteristics in auditory perception. The Journal of the Acoustical Society of America, 123, 366-376. https://doi. org/10.1121/1.2804951
Klatt, D. (1982). Prediction of perceived phonetic distance from critical-band spectra: A first step. ICASSP '82. IEEE International Conference on Acoustics, Speech, and Signal Processing, 1278-1281. https://doi.org/10.1109/ICASSP.1982.1171512
Kong, E. J., \& Edwards, J. (2011). Individual differences in speech perception: Evidence from visual analogue scaling and eyetracking. Proceedings of the International Congress of Phonetic

Sciences (ICPhS 17). Hong Kong, 17-21 August 2011 (pp. 1126-1129).
Kong, E. J., \& Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. Journal of Phonetics, 59, 40-57. https://doi.org/10.1016/j.wocn.2016.08.006
Morrison, G.S. (2007). Logistic regression modelling for first- and second-language perception data. In P. Prieto, J. Mascaró, \& M.J. Solé (Eds.), Segmental and prosodic issues in Romance phonology (219-236). Amsterdam: John Benjamins.
Morrison, G. S. (2009). L1-Spanish speakers' acquisition of the English /i/-/I/ contrast II: Perception of vowel inherent spectral change1. Language and Speech, 52(4), 437-462. https://doi. org/10.1177/0023830909336583
Morrison, G. S. (2013). Theories of vowel inherent spectral change. In G. Morrison \& P. Assmann (Eds.), Vowel inherent spectral change (pp. 31-47). Berlin-Heidelberg: Springer. https://doi. org/10.1007/978-3-642-14209-3_3
Nearey, T. M., \& Kiefte, M. A. (2003). A neural network approach to the dimensionality of the perceptual vowel space. Canadian Acoustics, 31, 16-17.
Pols, L. C. W., van der Kamp, L J. Th., \& Plomp, R. (1969). Perceptual and physical space of vowel sounds. The Journal of the Acoustical Society of America, 46, 458-467. https://doi. org/10.1121/1.1911711
Raizada, R. D. S., Tsao, F.M., Liu, H.M, \& Kuhl, P. K. (2010). Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: Prediction of individual differences. Cerebral Cortex, 20(1), 1-12. https:// doi.org/10.1093/cercor/bhp076
Rosner, B. S., \& Pickering, J. B. (1994). Vowel perception and production. Oxford, UK: Oxford University Press.
Shultz, A. A., Francis, A. L., \& Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. The Journal of the Acoustical Society of America, 132, EL95EL101. https://doi.org/10.1121/1.4736711
Stevens, K. N. (1998). Acoustic phonetics. Massachusetts: MIT Press.
Stevens, K. N., \& Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. The Journal of the Acoustical Society of America, 55, 653-659. https://doi. org/10.1121/1.1914578
Strange, W., \& Jenkins, J. J. (2013). Dynamic specification of coarticulated vowels. In G. Morrison \& P. Assmann (Eds.), Vowel inherent spectral change (pp. 87-115). Berlin-Heidelberg: Springer. https://doi.org/10.1007/978-3-642-14209-3_5
Werker, J. F., \& Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. Attention, Perception, \& Psychophysics, 37(1), 35-44. https://doi.org/10.3758/Bf03207136

