Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 4pSCb: Production and Perception I: Beyond the Speech Segment (Poster Session)

4pSCb16. The perception of formant-frequency range is affected by veridical and judged fundamental frequency

Santiago Barreda* and Terrance Nearey

***Corresponding author's address: University of Alberta, Edmonton, T6G 2E7, Alberta, Canada, sbarreda@ualberta.ca**

The vowels produced by different speakers vary in terms of their fundamental frequency (f0) and formant frequencies (FFs). Variation in the production of a given vowel category between speakers of different sizes is primarily according to a single multiplicative parameter (related to speaker vocal-tract length). This parameter, which we refer to as FF-scaling, has an associated perceptual quality that listeners may use to determine apparent speaker characteristics and vowel quality. In a previous experiment [Barreda & Nearey. 2011. J. Acoust. Soc. Am., 129, p. 2661], listeners were trained to identify a limited set of voices based on FF-scaling and f0 differences. The current study presented listeners with large number of voices (n = 4000) varying in FF-scaling and f0, arranged in a two-dimensional space where one dimension corresponded to each acoustic characteristic. Listeners were played a voice, and asked to indicate its location on the board, thereby providing an f0 and FF-scaling estimate for the voice. Results indicate that listeners are able to identify voice FF-scaling, and that this decision is informed primarily by veridical voice FF-scaling. However, there is a complicated relationship between perceived f0 and FF-scaling, suggesting an interdependent relationship in the perception of these characteristics.

Published by the Acoustical Society of America through the American Institute of Physics

Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

The range of formant frequencies (FFs) produced by a speaker are most strongly determined by the speaker's vocal-tract length, where speakers with longer vocal tracts produce lower FFs overall and speakers with shorter vocal tracts produce higher FFs overall (Fant, 1960). Vocal-tract length differences between speakers result in a situation in which two speakers may produce very different formant patterns when producing the same vowel category, just as they might produce very similar formant patterns to indicate two different intended vowel categories.

It has been suggested that differences in the oral/pharyngeal cavity ratios between adult males and females may result in non-uniform scaling of formant patterns between male and female speakers (Fant, 1975). However, no clear demonstration either of the statistical reliability of systematic non-uniformities in production, nor of the perceptual relevance of any such non-uniformities to listeners' identification performance exist in the literature. Furthermore, Turner et al. (2009) and Barreda and Nearey (2013) present good evidence that variation in the FFs produced by different speakers is mostly according to a single multiplicative parameter. For example, consider Figure 1a, which presents the Peterson & Barney (1952) vowel data. The line on this figure indicates a line parallel to log F1 = log F2, which indicates the direction of variation in FFs which can be explained according to a single multiplicative parameter. The major axes of the ellipses indicating different vowel categories are all closely aligned with the log F1 = log F2 line. An analysis presented in Barreda and Nearey (2013) indicates that roughly 80% of the variation in FFs between speakers occurs along this axis.

Variation according to a single multiplicative parameter entails that, if vowel quality is held constant, the FFs representing a given vowel sound when produced by one speaker will, on average, differ from those produced by another speaker by a single scalar value. We refer to this parameter as the formant-frequency scaling, or FF-scaling, associated with a voice, where voices with higher FF-scalings produce higher FFs overall and vice versa.

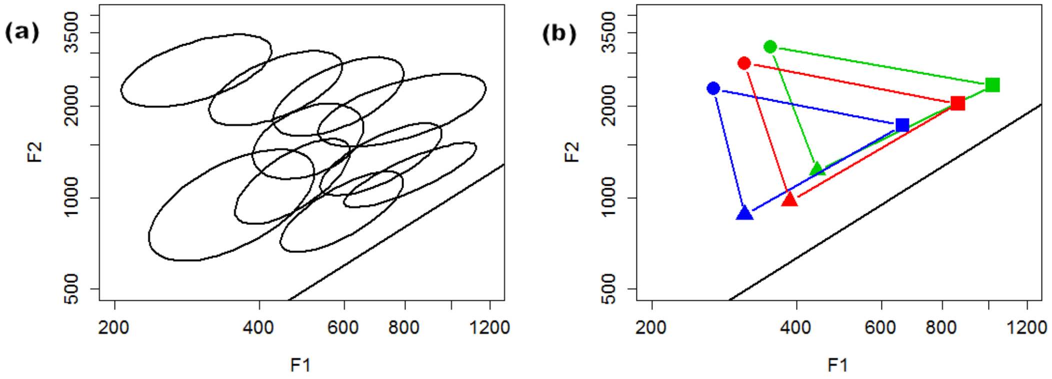

FIGURE 1. (a) Ellipses enclosing two standard deviations of the Peterson and Barney (1952) vowels are presented, where each ellipse indicates a different vowel category. (b) The mean location of /æ/ (square), /i/ (circle) and /u/ (triangle) are indicated for adult males (blue), adult females (red) and children (green) from the same data set. In both figures, the line is a line parallel to log F1 = log F2, which indicates the direction of variation according to a single multiplicative-parameter.

Figure 1b presents examples of the vowel spaces of speakers with different FF-scalings. For example, all three instances of /i/ represent instances of the 'same' vowel in that the linguistic content of the vowel sounds is fixed. One can imagine a situation in which speakers whose vowel spaces correspond to each of those indicated in the figure produce instances of /i/ at the same pitch (perhaps they belong to a choir). In this case, the vowels can be said to belong to the same category, and to have the same pitch, and yet they differ in terms of the FFs associated with them (i.e., they differ in terms of FF-scaling). Previous experiments have demonstrated that listeners show a sensitivity to FF-scaling, and that they use information related to this acoustic characteristic in making judgments of speaker height and gender (van Dommelen and Moxness, 1995; Rendell et al., 2007; Hillenbrand and Clark, 2009; Barreda and Nearey, 2012a). However, the ability of listeners to identify voice FF-scaling, or the mechanisms that might underlie this ability are not well understood.

In a previous experiment (Barreda and Nearey, 2013), we investigated the ability of listeners to report voice FF-scaling independently of f0. Listeners were presented with a board containing 15 buttons arranged in 3 rows of 5. Each of these buttons was associated with a stimulus voice with a unique combination of f0 and FF-scaling. Stimulus FF-scaling increased from left to right across columns, while f0 increased from top to bottom across rows (in a similar arrangement to that seen in Figure 2b). The lowest and highest f0 and FF-scaling levels had values appropriate for adult males and small children respectively. For each trial, listeners were presented with a stimulus voice and were asked to indicate its position on the board. In doing so, listeners were effectively providing an f0 and FF-scaling estimate for that voice.

Results indicated that listeners are able to identify the FF-scaling of voices with a good degree of accuracy after only a short training session. In addition, there was evidence to suggest that f0 interferences with the determination of FF-scaling so that judgments of voice FF-scaling may not solely be determined by the formant frequencies present in a stimulus sound. Finally, there was some evidence of a negative correlation between f0 and FF-scaling estimation errors so that f0 overestimations were associated with FF-scaling underestimations and vice versa.

Unfortunately, despite the positive results, the low number of stimulus voices and the relatively sparse sampling of the stimulus space did not allow for a detailed investigation into the association of f0 and FF-scaling identification errors, or the absolute accuracy with which listeners can identify voice FF-scaling.

The experiment to be described here is intended to address the shortcomings of Barreda and Nearey (2013), and to allow us to address some of the issues raised in that study. As in Barreda and Nearey (2013), listeners were presented with a stimulus voice with an unknown f0 and FF-scaling and were asked to indicate the location of the voice on a board. However, in this new experiment, 4000 unique stimulus voices (compared to only 15 in the earlier experiment) were densely packed onto the response space, with only very small acoustic differences between adjacent f0 and FF-scaling stimulus levels. Both of these changes were meant to simulate a continuous, rather than discrete, response space, and to allow for an investigation into the absolute accuracy of FF-scaling estimates and into the possible influence of f0 in FF-scaling identification.

METHODOLOGY

Participants

Listeners were 71 students from the University of Alberta drawn from a participant pool in which undergraduate students take part in experiments in exchange for partial course credit. All participants were students taking an introductory level, undergraduate linguistics course. Participants ranged in age from 17 to 25 and none reported any known hearing difficulty.

Stimuli

The stimuli consisted of synthetic vowel pairs with formant-patterns appropriate for the vowels [i æ] as spoken by a range of synthetic voices. These voices varied on the basis of FF-scaling and average f0. Voices were created at 100 f0 levels, fully-crossed with 40 FF-scaling levels, resulting in 4000 unique stimulus voices. The frequencies of the first four formants for the vowels representing the lowest FF-scaling level are provided in Table 1. Formants above F4 were set at 1000 Hz higher than the previous formant for the lowest FF-scaling level. Vowels were always presented in the same order [i æ], and each vowel was 200 ms in duration. The two vowels were separated by 150 ms of silence in a single sound file synthesized at a sampling frequency of 22050 Hz.

TABLE 1. Formant frequencies (in Hz) for the stimulus vowels representing voices at the lowest FF-scaling level.

Vowel	F1	F2	F3	F4
i	275	2114	2711	3500
æ	705	1473	2281	3500

Voices which differ only in terms of FF-scaling will produce vowels of the same category whose FFs differ, on average, by a single multiplicative parameter (Nearey, 1978; Nearey and Assmann, 2007; Turner et al., 2009;

Barreda and Nearey, 2013). As a result, the FFs used to represent the voice with the lowest FF-scaling may be scaled up by a given percentage to simulate the vowels of speakers with higher FF-scalings. To create a series of vowels which differ in terms of FF-scaling, a scalar value was associated with each FF-scaling level. The FFs presented in Table 1 were then multiplied by this scalar value in order to determine the FFs for each FF-scaling level.

Appropriate FF-scaling ranges were determined with respect to the /i/ produced by each voice. In the Peterson & Barney (1952) data, the range of the average F1-F2-F3 frequencies for /i/ across all speakers is 1693 to 2807 Hz. In the Hillenbrand et al. (1995) data, the range for this value is 1698 to 2683 Hz. In accordance with these naturally observed ranges, the highest and lowest FF-scaling levels were set so that the mean of the first three FFs of /i/ would equal 1700 Hz for the lowest FF-scaling level and 2700 Hz for the highest FF-scaling level. Intermediate FF-scaling levels were interpolated between these two extremes in equal logarithmic steps. This resulted in an increase of approximately 1.2% between adjacent FF-scaling steps and a total increase of 58% in FF-scaling between the highest and lowest levels.

The f0 levels used for stimulus voices were determined in terms of the f0 at the midpoint of the vowel. Midpoint f0 levels spanned from 100 to 300 Hz in 100 equal logarithmic-steps. This means that each f0 level was roughly 1.1% higher than the f0 of the previous level. For each vowel, f0 decreased linearly from the beginning to the end of the vowel from 10% higher than midpoint f0, to 10% lower than the midpoint f0.

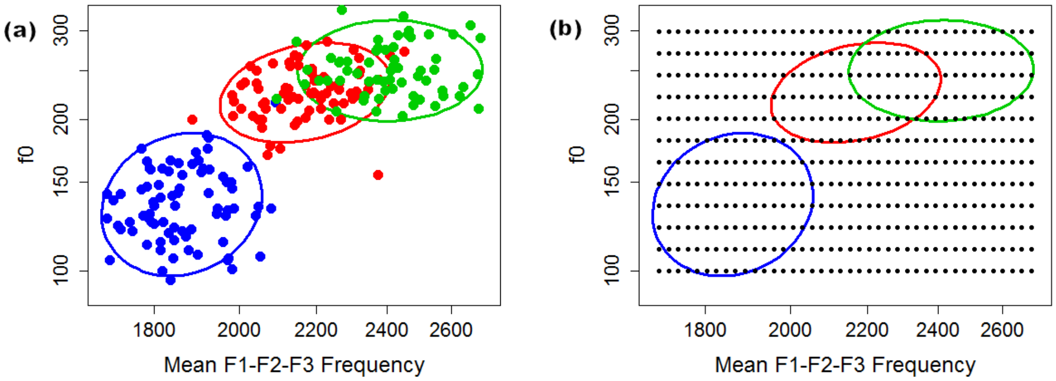

FIGURE 2. In both figures, the x-axis indicates the mean of the first three formant frequencies for productions of /i/. Ellipses enclose two standard deviations of the distribution of real voices from data collected by Peterson and Barney (1952) and Hillenbrand et al. (1995), representing data collected from 215 speakers. Ellipses indicate the distribution of voices of adult males (blue), adult females (red), and children (red). (a) The points indicate the locations of the voices of individual adult males (blue), adult females (red) and children of either gender (green) from the aforementioned data sets. (b) The locations of stimulus voices are indicated by the filled points. To maintain the legibility of the figure, only every 9th f0 level is indicated.

Figure 2b presents a comparison of the stimulus voices with the characteristics of natural data reported in two large data sets. The stimulus voices can be said to cover the essentially the entire typical range of f0 and FF-scaling observed for natural voices. However, since f0 and FF-scaling stimulus levels were fully crossed, stimulus voices encompass both typical f0 and FF-scaling combinations (the bottom-left and top-right quadrants of Figure 2a), and atypical combinations of these acoustic characteristics (the top-left and bottom-right quadrants of the figure).

Procedure

Listeners were presented with a graphical user interface displaying a 900 by 700 pixel board. Each of the 4000 stimulus voices was associated with a specific pixel location on this board. Stimulus voices were arranged so that FF-scaling increased left to right, and f0 increased top to bottom. In fact, the stimulus voices were arranged on the board in the same manner as shown in Figure 2. Adjacent FF-scaling levels were separated by 20 pixels along the horizontal axis, while adjacent f0 steps were separated by 6 pixels along the vertical axis.

The points associated with the stimulus voices only spanned a 780 by 594 pixel subsection of the response board. This subsection was centered so that there was a 60 pixel buffer around the most extreme stimulus voices on the horizontal axis, and a 53 pixel buffer around the most extreme voices on the vertical axis. When listeners clicked on an area of the response space not directly associated with any stimulus voice, the closest stimulus voice was played.

The interface were designed so that listeners could make mistakes in all directions for all stimulus voices, and so that they would not be directly aware of where the stimulus voices ceased to vary near the edges of the board. These aspects, combined with the very small acoustic differences between stimuli, were intended to encourage listeners to treat the response board as a continuous space.

Before beginning, listeners were given an opportunity to click on the board 25 times. Each time they clicked, the stimulus voice nearest to their click was played to them, and a blue circle was placed on the board. This was done to give listeners a rough idea of what voices in different locations on the board sounded like before beginning the experiment proper.

The procedure for the experimental task was as follows. A stimulus voice was selected at random, with the only constraint being that it be at least 4 FF-scaling steps and 10 f0 steps away from the previous stimulus voice. The listener was presented with this stimulus voice over headphones in a sound-attenuated booth. They were given the opportunity to replay this voice up to 4 times before providing any responses. To provide responses, the listener was asked to click on the location of the board they felt was associated with the stimulus voice they had just heard. When the listener provided a response, the location of their click was indicated by a blue circle centered at this location. The stimulus voice whose location was nearest to the user's mouse click location was then played for the user.

Depending on the block, listeners may have been asked to provide more than one response. In that event, after providing a first response, listeners could provide further responses by clicking on the board again. Each time a listener provided a guess, the location of their guess was indicated by a blue circle and the stimulus voice located closest to their click was played. Listeners thus had the opportunity to compare the sound of the location they picked with their memory of how the test stimulus sounded.

Blocks differed on the basis of the number of times the listener was asked to guess the location of each presented stimulus. In the first block, listeners were given three opportunities to respond. In the second block, listeners were given two opportunities to respond, and in the final block, only a single response was collected.

When the listener had completed the appropriate number of responses given the block, the correct location of the stimulus voice they had heard was indicated by a red circle centered about the coordinate associated with the stimulus voice. A bull's-eye was drawn around the red circle to give participants some indication of how close their guesses had been. After a 1 second pause, the next stimulus played automatically and the process repeated itself until the block was completed. Since listeners were free to proceed at their own pace, the blocks were given time limits and listeners completed as many trials as they were able within the allotted time. Listeners were allotted 15 minutes for the first block (3 responses), and 12 minutes each for the second (2 responses) and third (1 response) blocks.

RESULTS

We will limit our discussion to the results of the third block, where listeners provided only a single response for each stimulus voice. Listeners provided an average of 126 responses, with a minimum of 51 and a minimum of 185 responses per listener, for a total of 4254 responses across all listeners.

Of primary interest was the size of average FF-scaling identification errors, and the relationship between these errors and errors in f0 identification. In both cases, errors are defined as the distance, measured in pixels, between the location on the response board associated with a stimulus voice, and the location indicated for a speaker during a trial. Since stimulus voices were placed on the board at fixed intervals (measured in pixels), and these intervals were associated with fixed differences in f0 and FF-scaling (measured in Hz), error distances measured in pixels may be converted to hertz in a straightforward manner.

Since stimulus step sizes were created using a logarithmic scale, an error of a given number of pixels is associated with an error of a given percent relative the acoustic properties of the test stimulus, regardless of the position on the board. For example, since a difference of 20 pixels along the x-axis is associated with an increase of 1.2% to all FFs, a 20 pixel overestimation for a stimulus with average FFs of 1800 Hz means that a listener indicated average FFs of 1822 Hz (1800 x 1.012), while the same error for a stimulus voice with average FFs of 2400 Hz meant that a listener indicated average FFs of 2429 Hz (2400 x 1.012). As a result of this, errors will be discussed in terms of errors as a percentage of the actual acoustic characteristics of the test stimulus, where an error of a single pixel is associated with an error of 0.059% in FF-scaling and 0.185% in f0. Since adjacent f0 steps are closer together on the board than adjacent FF-scaling steps, the magnitude of f0 errors may be artificially inflated relative to the magnitude of FF-scaling errors. It is important to keep in mind that the objective of the present study is not to compare the magnitude of f0 and FF-scaling errors, but to investigate the nature of FF-scaling identification specifically.

The experimental design was intended to result in an approximately continuous response space in the hope that errors in f0 and FF-scaling identification would follow a bivariate normal distribution. As seen in Figure 3, this was successful: f0 and FF-scaling errors are normally distributed, and the joint distribution of errors can reasonably be treated as bivariate normal.

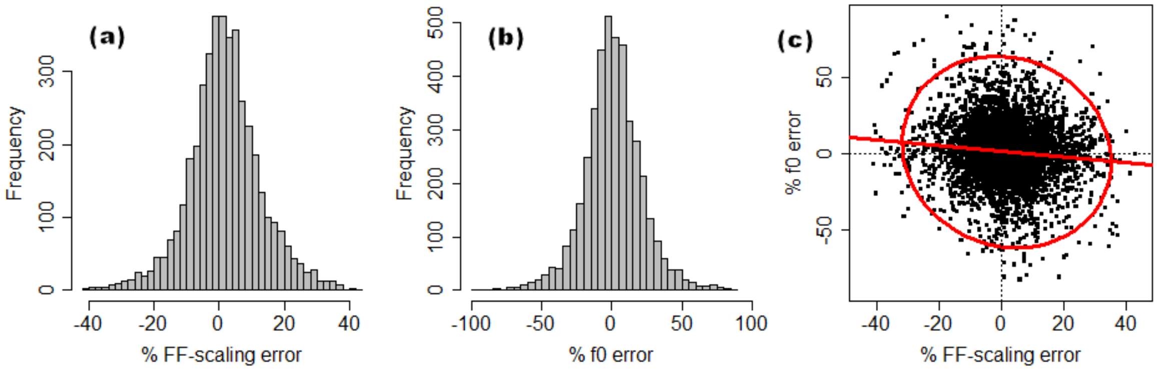

FIGURE 3. (a) FF-scaling errors pooled across all listeners. (b) f0 errors pooled across all listeners. (c) Each point indicates the location of an individual error, pooled across all listeners. The ellipse encloses 3 standard deviations of pooled errors. The red line indicates the line of best fit relating FF-scaling errors and f0 errors, pooled across all participants.

The magnitude of the average absolute FF-scaling error across listeners was 8.3% (min = 5.4%, max = 12.1%). The magnitude of the average absolute f0 error across listeners was 15.6% (min = 10.9%, max = 27.3%). The average within-participant FF-scaling error was 1.74% [t(33) = 4.7, p < 0.0001], meaning there was a small, but statistically-significant tendency for listeners to overestimate stimulus FF-scaling. The average f0 error was 0.94%, suggesting a slight inclination towards overestimation of stimulus f0, however, this did not reach statistical significance [t(33) = 1.95, p = 0.0595].

A visual inspection of Figure 3c suggests that errors in judged f0 and FF-scaling may in fact be negatively correlated. To investigate this, the correlation coefficient between f0 and FF-scaling errors was found for each listener, revealing an average, within-participant correlation of -0.09 [t(33) = -3.4, p = 0.0017] between f0 and FF-scaling errors.

Information used in FF-scaling estimation

The x and y axis coordinates representing the listener's response indicate judged FF-scaling and f0 respectively. Of primary interest to our investigation is judged FF-scaling, which is the x-axis coordinate associated with the listener's response. If listeners used only information related to stimulus FF-scaling when determining judged FF-scaling, there should be no significant role for any information related to stimulus f0. If stimulus f0 affects judged FF-scaling, there are two additional variables which might affect judged FF-scaling. The first is stimulus f0, which may be indexed using the y-axis coordinate associated with the stimulus voice for a given trial. The second is f0 error, which is the difference between the y-axis coordinate indicated by the listener and the actual stimulus y-axis location, where positive values indicated overestimations and negative values indicate underestimations.

To investigate what information plays a role in FF-scaling estimation, a random coefficients regression model was carried out (Gumpertz and Pantula, 1989). Regression analyses were carried out on the responses collected from each participant, and significance testing was carried out on the estimated coefficients across all listeners. For each model, judged FF-scaling (i.e., the x-axis coordinate indicated by the listener for a trial) was the dependent variable. The independent variables were the actual stimulus FF-scaling coordinate, the actual stimulus f0 coordinate, and the error in identification of stimulus f0. All values were coded in terms of locations or distances (in pixels) on the response board.

The results of this analysis indicate that stimulus FF-scaling [t(33) = 25, p < 0.0001], stimulus f0 [t(33) = -3.3, p = 0.002] and f0 error [t(33) = -5.1, p < 0.0001] all have a significant effect on judged FF-scaling. Stimulus FF-scaling had a positive effect on judged FF-scaling, while both judged f0 and f0 error had negative effects. To get an idea of the relative importance of these predictors, the same model was fit to the pooled data across all listeners.

This analysis revealed that stimulus FF-scaling explains 44.6% of the variance in judged FF-scaling, while judged f0 and f0 error explain only 0.52% and 0.89% respectively.

DISCUSSION

The just noticeable difference for FF-scaling has been estimated to be 7-8% for isolated vowels by Smith et al. (2005) and 4-6% for syllables by Ives et al. (2005). In both cases, just noticeable differences were estimated using a two-alternative, forced-choice methodology. In this experiment, the average FF-scaling error was 8.3% across all listeners, and the average error of some listeners was as small as 5%. Given that every listener performed at least 50 trials, this is not simply a case of high performance due to chance over a small number of trials. Furthermore, in this experiment, listeners had to identify voices that spanned a very large range of FF-scalings, and had to simultaneously identify voice f0, which was fully crossed with FF-scaling levels. Given all this, listeners show a remarkable ability to identify voice FF-scaling

The analysis presented in the previous section indicates that FF-scaling judgments are influenced strongly by the FFs associated with a voice, and weakly affected by stimulus f0. However, the direction of the effects for stimulus f0 and f0 error are in the opposite direction that one might expect given the natural covariation between f0 and FF-scaling, where speakers with higher f0s also tend to produce higher FFs overall (Nearey and Assmann, 2007). If listeners were simply using this covariation to guess stimulus FF-scaling, we would expect that higher stimulus f0, and in particular larger f0 overestimations, would result in higher judged FF-scaling responses.

One possible explanation for this behaviour is that listeners may sometimes "work backwards" from some internal speaker size estimate to identify stimulus FF-scaling and f0 values. Many experiments have demonstrated that listeners can readily make and report speaker size judgments on the basis of stimulus f0 and FF-scaling properties (van Dommelen and Moxness 1995, Rendell et al. 2007, Hillenbrand and Clark, 2009, Barreda and Nearey, 2012a). If listeners find it easier to identify speaker size than stimulus f0 and FF-scaling, they may be using apparent speaker size as a constraint on possible stimulus properties. For example, consider a given stimulus with a fixed f0 and FF-scaling level. Consider a case in which the listener is reasonably certain that this speaker is a small child approximately 1 meter in height. If the listener misattributes some of this apparent ‘smallness’ to f0 by overestimating f0 (resulting in positive f0 error), they must then underestimate stimulus FF-scaling in order to maintain their current estimate of the speaker’s height.

This version of events would suggest that listeners may have some difficulty in reporting stimulus FF-scaling and, in some cases, may use apparent speaker characteristics in order to identify this property. However, it is important to keep in mind that, although significant, the effect for stimulus f0 and f0 errors on judged FF-scaling were quite weak, and listeners were quite accurate in identifying stimulus FF-scaling overall. Taken together, these facts suggest that the strategy outlined above may be a secondary strategy employed by listeners in cases of doubt, or may be used by listeners to fine-tune their FF-scaling estimates.

CONCLUSION

In a previous experiment (Barreda and Nearey, 2013), we found that FF-scaling judgments are influenced by stimulus f0 in addition to being informed by stimulus FF-scaling. We also found evidence to suggest that errors in FF-scaling and f0 judgments were negatively correlated so that overestimation of one characteristics was associated with an underestimation of the other characteristic. The results presented here confirm these findings. Listeners are able to report voice FF-scaling with a high level of accuracy, and these judgments are influenced by stimulus f0. Furthermore, our results replicate the finding, reported in Barreda and Nearey (2013), that errors in f0 and FF-scaling judgments are negatively correlated.

It is important to note that, in this experiment, the phonetic identity of stimulus sounds was fixed, and the vowel pair presented to listeners spanned the entire F1 range for each stimulus voice. In cases of more phonetic ambiguity, when both speaker identity and vowel quality vary unpredictably from trial to trial, the relation between f0 and FF estimates might be different. For example, in the experiment described in Barreda and Nearey (2012b), listeners were presented with vowel sounds at different f0 and FF-scaling levels and were asked to report phonetic identity and FF-scaling. In that experiment, we found that increasing stimulus f0 levels led to higher reported FF-scalings.

In the future, experiments investigating the relationship between FF-scaling estimation and perceived vowel quality will need be carried out in order to more closely investigate the relationship between FFs, f0, vowel quality and FF-scaling in cases where apparent speaker characteristics and vowel quality are unknown to the listener.

REFERENCES

Barreda, S. and T.M. Nearey. (2012a). The direct and indirect roles of fundamental frequency in vowel perception. Journal of the Acoustical Society of America 131: 466-477.

Barreda, S. and Nearey, T. (2012b). The association between speaker-dependent formant space estimates and perceived vowel quality. Canadian Acoustics 40: 12-13.

Barreda, S. and T.M. Nearey. (2013). Training listener to report the acoustic correlate of formant-frequency scaling using synthetic voices. Journal of the Acoustical Society of America 133(2). To appear.

Fant, G. (1960). Acoustic Theory of Speech Production. The Hague: Mouton. pp.107-138.

Fant, G. (1975) Non-uniform vowel normalization, STL-QPSR 2-3: 1 – 19.

Gumpertz, M., and Pantula, S. G. (1989). A Simple Approach to Inference in Random Coefficient Models. The American Statistician, 43(4), 203-210. doi:10.2307/2685362

Hillenbrand, J.M., Getty, L.A., Clark, M.J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America, 97, 3099-3111.

Hillenbrand, J.M., and Clark, M.J. (2009). The role of F0 and formant frequencies in distinguishing the voices of men and women. Attention, Perception, and Psychophysics, 71, 1150-1166.

Ives, D. T., Smith, D. R. R. and R. D. Patterson. (2005). Discrimination of speaker size from syllable phrases. Journal of the Acoustical Society of America 118: 3816-3822.

Nearey, T. M. (1978). Phonetic Feature Systems for Vowels. PhD thesis, Indiana University Linguistics Club.

Nearey, T. M. and Assmann, P. F. (2007). Probabilistic "sliding template" models for indirect vowel normalization. In Maria-Josep Solé, Patrice Beddor, and Manjari Ohala (eds.) Experimental Approaches to Phonology. Oxford: Oxford University Press. 246-69.

Peterson, G. E. and Barney, H. L. (1952). Control methods used in a study of the vowels. Journal of the Acoustical Society of America 24: 175-184.

Rendall, D., Vokey, J. R., and Nemeth, C.. (2007). Lifting the Curtain on the Wizard of Oz: Biased Voice-Based Impressions of Speaker Size. Journal of Experimental Psychology: Human Perception and Performance 33: 1208 –1219.

Smith, D. R. R., Patterson, R. D., Turner, R, Kawahara, H. and T. Irino. (2005). The processing and perception of size information in speech sounds. Journal of the Acoustical Society of America 117: 305-318.

Turner, R. E., Walters, T.C., Monaghan, J., and Roy D. Patterson. (2009). A Statistical, Formant-pattern Model for Segregating Vowel Type and Vocal-tract Length in Developmental Formant Data. Journal of the Acoustical Society of America 125: 2374-2386.

van Dommelen, W. A. and Moxness, B. H. (1995). Acoustic Parameters in Speaker Height and Weight Identification: Sex-Specific Behaviour. Language and Speech 38: 267-287.