HOW THE ROLE OF F3 IN VOWEL PERCEPTION MAY BE INFLUENCED BY LISTENER EXPECTATIONS

Santiago Barreda, and Terrance M. Nearey
Dept. of Linguistics, University of Alberta, 4-32 Assiniboia Hall, Edmonton, AB T6G 2E7, Canada sbarreda@ualberta.ca

1. INTRODUCTION

The relationships among apparent speaker identity, a vowel's physical properties and its phonetic quality are not fully known. In a previous experiment, Experiment 1, we studied the relationship between f0, apparent speaker identity and vowel quality by asking participants to make simultaneous vowel and apparent speaker size and gender judgments. We considered three possibilities: that f0 might be directly related to vowel quality in the same way that the formants are, that f0 mainly affects vowel quality indirectly by affecting apparent speaker characteristics, and that f0 is not related to vowel quality at all. Results indicated that f0 affects vowel quality mainly indirectly, via its effects on apparent speaker characteristics. However, some listeners find judgments of size ('large' to 'small') difficult to make consistently and listeners may adapt different size criteria within and between genders. Experiment 2 was similar to Experiment 1, but listeners were instructed that all the synthetic voices were males and that they should report speaker's apparent age. We compare the results of the two experiments that involve the same stimuli and same phonetic responses, but different instructions and 'size-related' responses. There are some surprising results that suggest listeners' use of F3 and higher formants in forming phonetic judgment varies with differences in simultaneous non-phonetic tasks.

2. METHOD

2.1 Participants and stimuli

Participants were 25 students from the University of Alberta. All participants were students taking an introductory level, undergraduate linguistics course. Participants were drawn from a participant pool in which undergraduate students take part in experiments in exchange for partial course credit.

A continuum was designed that spanned roughly from the average F1-F2 frequencies of the /ʌ/ of a male to those of the average /æ/ produced by a female in seven equally spaced steps. The fourth point of this continuum had F1-F2 frequencies appropriate for either an /æ/ produced by an adult male or an /ʌ/ produced by an adult female. Production data collected at the Alberta Phonetics Lab indicated that F3 distributions were nearly identical for the two vowels, implying that F3 could carry little to no direct phonetic information. It was therefore expected that F3 could be manipulated without directly affecting the phonetic quality of the vowels. This seven-step F1-F2 continuum was combined with three different higher formant conditions (where higher formants F4 through F10 varied proportionally with F3) and three different f0 conditions. This yielded a total of 63 distinct vowel stimuli. The frequencies of all of the continuum points and f0 and F3 levels used are presented in Table 1.

	f0 Levels			F3 Levels		
	L	**M**	**H**	**L**	**M**	**H**
μ	108	153	215	2475	2755	3068

#	**1**	**2**	**3**	**4**	**5**	**6**	**7**
F1	684	735	789	848	911	978	1051
F2	1354	1455	1563	1679	1803	1937	2081

Table 1. Formant frequencies and f0 levels for the stimulus vowels used.

2.3 Procedure

2.3.1 Instructions and judgments.

The same stimuli were used in two experiments. In both experiments, participants were instructed that they would be hearing a human-like, 'robotic' voice producing vowels intended to be either /ʌ/ or /æ/. Participants were asked to listen to the vowel and decide which of the two vowel categories the vowel sounded most like.

In Experiment 1, 19 listeners were told that the speakers were either males or females of varying sizes. In addition to the vowel responses, they were required report the gender and relative size of the apparent speaker, by clicking on appropriately marked response areas. (Barreda & Nearey 2010). In the new Experiment 2, listeners were told that these speakers were all male but varied in age from 5 to 25 years of age. Listeners were asked to indicate vowel quality the apparent age of the speaker. Typically, listeners responded to three repetitions of the stimulus list (189 responses), followed by a short break, after which the participant performed another three repetitions of the same list. A total of 9,450 responses were collected across all 25 participants.

2.3.2 Comparison of the two experiments

Experiments 1 and 2 were identical except for (a) the different instructions given to the participants about the

synthetic voices and (b) the non-phonetic response variables. We focus here on comparing only those instances from Experiment 1 in which participants indicated that they thought they were listening to a male speaker. This included 4,548 responses used out of a total of 6,921 responses overall. Since the vowel stimuli were identical in both experiments and only trials where participants had heard a male speaker were considered, any significant differences across both experiments must be attributable to the different instructions given or to differences induced by the differences in the non-phonetic tasks.

3. RESULTS

Initially, the objective of this second experiment was to see if apparent speaker age explained more of the variance in vowel quality than apparent speaker size. Results indicated that age in Experiment 2 was a worse response variable than speaker size in Experiment 1 in terms of its correlation with vowel quality. To further investigate the differences between the two experiments, logistic regression was carried out within-participant using vowel category as a response variable and F1, F3 and f0 as the predictor variables. The estimated coefficients were collected across all participants and compared across both experiments.

	F1	F3	f0
Experiment 1	40.22	-16.15	-2.72
Experiment 2	34.26	-4.25	-2.94

Table 2. Means of the estimated within-participant coefficients across both experiments.

Two sample t-tests were carried out on the within-participant coefficients across both experiments to see if the relationships between vowel quality and each of the three predictors (F1, F3, f0) differed between the two experiments. For example, the mean effect for F1 was 40.22 in the first experiment and 34.26 in the second experiment but this difference was not statistically significant ($t = 1.58$, $df = 41.26$, $p = 0.123$). Nor was there a significant difference for the f0 coefficients ($t = 0.3$, $df = 30.48$, $p = 0.767$). However the difference between the estimated F3 coefficients is highly significant ($t = -5.28$, $df = 22.13$, $p < .0001$), with Experiment I showing a substantially larger effect of F3.

4. DISCUSSION

Our results indicate that F3 has a substantially weaker effect on vowel decisions in Experiment 2 than Experiment 1. Since the stimuli in the two experiments were the same, the only differences between the two experiments are in the instructions given to participants and in the responses used to gauge apparent speaker characteristics.

The most striking result in this experiment, the differential weight of F3, was a surprise, for which we have no compelling explanation. Instead we will speculate on two possible factors that we hope to explore in future experiments.

One factor is that the variety of speakers contemplated by listeners was greater in Experiment 1. Furthermore, the gender distinction is a categorical one that seems to be very salient for listeners. It is possible that this induced listeners to pay more attention to F3 as a potential cue to speaker identity. Even though we are considering only the vowel categorization for voices judged as male, the additional attentiveness to F3 may have carried over to within-gender comparisons.

The second factor is that in Experiment 2, voices were described as male speakers only and the only size-related judgment, age, was a continuous one. Listeners reported some difficulty in estimating ages. Perhaps because of this uncertainty, they were generally less likely to make size-related voice distinctions of the kind we hoped to induce by varying F3.

To further investigate these ideas, we will need to conduct additional experiments, such as combining age and gender judgments, or asking for judgments of size without gender. In any event, the differential use of F3 in judgment of vowel quality is a phenomenon that deserves more thorough study. Our results also indicate that there are interactions between F3 and f0 which make the effect of F3 on vowel quality weaker as f0 increases.

Most of the theories of vowel perception we have encountered do not readily accommodate such variable cue weighting. Additional experiments and theory refinement are clearly required.

REFERENCES

Barreda, S. & T.M. Nearey. (2010). The relationship between fundamental frequency and vowel quality. *JASA, vol. 127, issue 3, p. 2019*

ACKNOWLEDGEMENTS

This work was supported by funding from SSHRC and carried out at the facilities at the Alberta Phonetics Lab and the Center for Comparative Psycholinguistics at the University of Alberta.