Training listeners to report fundamental frequency and formant range information independently

Santiago Barreda & Terrance M. Nearey

Alberta Phonetics Laboratory

Social Sciences and Humanities Research Council of Canada | Conseil de recherches en sciences humaines du Canada | Canada

Introduction

A listener's impression of the apparent speaker is informed primarily by the fundamental frequency (f0) and average formant range (FR) of the voice.

When investigating the relationship between speech perception and the apparent speaker, collecting speaker characteristics which are only partly determined by a voice's formant frequencies (FFs) is an indirect way of obtaining information regarding a listener's estimation of that voice's FR.

If listeners could report estimated f0 and FR information separately, the independent effect of either of these estimates on vowel perception could be investigated.

Objectives: 1) To develop a training procedure during which listeners learn to report the f0 and FR of a voice independently. 2) To see how well listeners can learn to make these distinctions. 3) To see how small an FF difference listeners can learn to identify.

Methods

Participants: 71 listeners. Each listener was randomly assigned to each one of four scale factor groups.

Stimuli: Instances of / i æ / from 15 synthetic 'voices' which differed from each other in their FRs and/or their f0s. Voices' FRs differed from each other by a constant scale factor. Four scale factors were used in all (7%, 8%, 9%, 10%). Scale factor was a between-subjects factor.

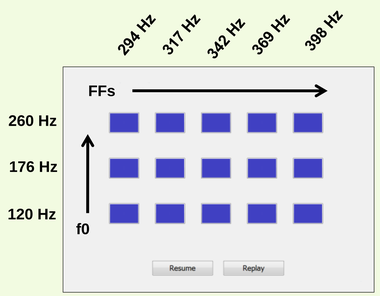


Figure 1 *– A screenshot of the game board used. Each button was associated with a single voice. Voices differed in f0 across rows and in their FR across columns. The F1 value for / i / at the 8% scaling level is given as an example of FR increases. f0 levels shown were constant across all scaling levels.*

Procedure: Participants were presented with a computer interface consisting of a board with a series of buttons on it. Each button was always associated with the same voice.

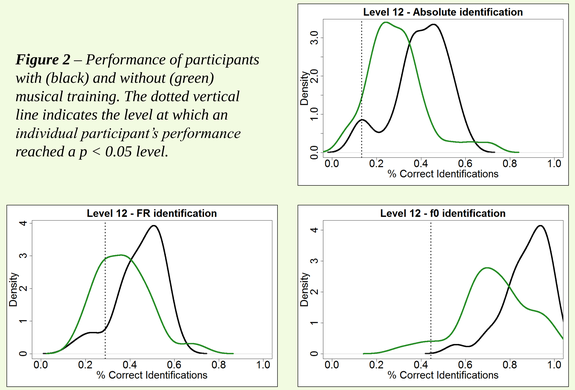
Participants progressed through a series of levels in which the number of choices increases from level to level. In each level, participants heard a voice and had to select it from a set of available choices by clicking on the correct button. Voices that were available during a trial were indicated by a blue button. Participants were given feedback when they made an error.

Results

When participants had to choose from among all voices and report exact f0 and FR. This was the penultimate level and the first in which listeners had to identify a voice from among all possible voices.

- **Overall listeners identified 33% of voices correctly**. This is excellent performance given that we used FF differences as small as 7% and [1] estimated the JND for FF changes to be 8%.
- **Musically trained participants scored an average of 11% higher** [t(70) = 3.55, df = 70, p = 0.0007] than those without musical training.
- **Listeners identified 79% of voice f0s correctly**. Musically trained listeners scored an average of 14% higher [t(70) = 2.8, p = 0.0064]
- **Listeners identified 40% of voice FRs correctly**. Musically trained listeners scored an average of 8% higher [t(70) = 4.4, p = 0.00004].

Figure 2 *– Performance of participants with (black) and without (green) musical training. The dotted vertical line indicates the level at which an individual participant's performance reached a $p < 0.05$ level.*



Analysis of Errors.

- Listeners scored an average of 7% more [t(46)= 3.1, p = 0.003] correct FR identifications when they correctly identified a voice's f0. Only participants who made at least 5 f0 errors were considered.
- When listeners made both FR and f0 errors, **there was a strong tendency to trade-off FR overestimations with f0 underestimations and vice-versa** (see Table I).

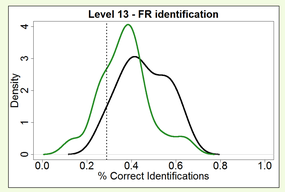
Table 1 *– Instances in which participants made both f0 and FR errors, pooled across all listeners. There were 450 errors of this kind and an expectation of 125 per cell. Only 33 listeners are included in this table.*

	Underestimate f0	Overestimate f0
Underestimate FR	91	161
Overestimate FR	129	69

When participants had to choose from among all voices and report exact FR and disregard f0. This was the final level and all voices were again available. Listeners had to identify a voice's FR only an disregard f0 information. Participants indicated a voice's FR by clicking on the correct column in the middle row of the board.

- Overall listeners identified 40% of voices correctly. **There was no difference between FR identification rates for levels 12 and 13**.
- **Musically trained participants identified an average of 8% higher** [t(70) = 3.16, p = 0.002] than those without musical training.

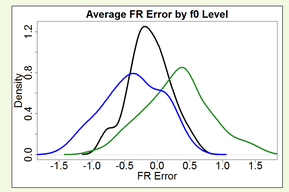
Figure 3 *– Performance of participants with (black) and without (green) musical training. The dotted vertical line indicates the level at which an individual participant's performance reached a $p < 0.05$ level.*



Analysis of Errors.

- In level 13**, listeners were more likely to overestimate FR when the voice had a low f0 and to underestimate FR when the voice had a low f0** (see Figure 4).
- On average, listeners overestimated FR by .3 steps on low f0 voices, underestimated FR by .12 steps on mid f0 voices and underestimated FR by -.36 on high f0 voices.
- A repeated-measures ANOVA reveals that this effect is highly significant [F(1,209) = 16.7, p = 0.00006].

Figure 4 *– Kernel density estimates for average, within-participant FR errors when the voice had low (green), mid (black) and high (blue) f0.*



Conclusion

Listeners ARE able to dissociate f0 and FR information to a good degree.

- [1] estimated a JND of 8% for changes in FFs. By the end of this training, listeners were able to absolutely identify voices with FF differences as low as 7% between them.

There was an association between FR and f0 errors.

- When listeners made both f0 and FR errors, these errors were negatively correlated.
- When listeners were presented with high f0 voices, they were more likely to underestimate FR. When listeners were presented with low f0 voices, they were more likely to overestimate FR.

Listeners with musical training showed an advantage in every level and in both f0 and FR identifications.

- In the future, we will keep track of musical training to see if it has any effects on other tasks involving listener judgments.

References

[1] Smith, D.R.R., Patterson, R.D, & R. Turner . (2005).The processing and perception of size information in speech sounds. *J. Acoust. Soc. Am.* 117: 305-318.

Contact Information

Santiago Barreda (sbarreda@ualberta.ca)

Terrance M. Nearey (t.nearey@ualberta.ca)

Department of Linguistics, 4-32 Assiniboia Hall, University of Alberta, Edmonton, Alberta, Canada T6G 2E7