Individual Differences in F0 Imitation: Causes and Effects

M. Postma, E. Postma

TiCC, Tilburg University, Tilburg, The Netherlands
m.nilsenova@tilburguniversity.edu; e.o.postma@tilburguniversity.edu

Introduction

Phonetic imitation plays an important role in human interaction in that it reflects the closeness of the social bond between two individuals. Past studies have indicated the importance of the region between 50 Hz to 300 Hz (the fundamental frequency (F0) region) which is the most important source of information regarding emotions, stands and attitudes in the voice [1,2]. The same region also provides acoustic information for imitation exploited in promoting social convergence and status accommodation [3-8] and expressing ingroup-outgroup bias [9]. Interestingly, there appear to be large individual differences in speakers' ability to imitate F0. Using the shadowing task paradigm, originally introduced by [10], a recent study [11] found a considerable amount of variation in F0 accommodation, with some subjects actually diverging from the F0 of the model talker.

We hypothesized that the individual differences in speakers' ability to imitate F0 may at least partly be due to their neurocognitive ability to extract information about pitch from the speech signal. In particular, due to neuroanatomical differences found in the lateral Heschl's gyrus (the 'pitch processing center'), some listeners show an auditory perception bias for the sound as a whole (fundamental listeners), while others (spectral listeners) focus on its harmonic constituents [12,13]. The auditory perception bias has been almost exclusively analyzed in the context of musical training, but the results of individual studies indicate that it may also affect linguistic performance [14,15]. This study is the first attempt to explore the role of perception bias in imitation and thus its possible impact on social convergence.

Experimental study

Participants. Eighty-eight Dutch native speakers (67 female) between the age of 17-25y (M=20.48, SD=2.12) participated in the experiment for course credit. None of them reported any hearing difficulties. Fourteen of the participants were left-handed; about a half of the experimental group described their musical proficiency as low to average, the other half assessed their proficiency as high to professional.

Measuring auditory perception bias. Participants' auditory perception bias was determined with a variation of the psychoacoustic perceptual test described in [16], [17] and [18]. For the perceptual test, we constructed 36 pairs of complex harmonic tones, all 160 ms long, that consisted of 2-4 harmonics, with the same harmonic composition as employed by [17]. Participants were asked to categorize 18 perceptually ambiguous stimuli consisting of two complex tones, A and B, that were composed of a number of upper harmonic tones with the same highest harmonic but different levels of virtual fundamental pitch (derived from the harmonics as the best fit) and spectral pitch (based on the lowest harmonic). For example, the sequence A = (800 Hz + 1000 Hz) and B = (667 Hz + 1000 Hz) would be perceived as rising by a fundamental listener who would derive the missing F0 to be 200 Hz for tone A and 333 Hz for tone B; it would, however, be perceived as falling by a spectral listener who would focus on the lowest harmonics, i.e., 800 Hz and 667 Hz. The other 18 stimuli served as control trials in that their interpretation is unambiguous but helps to determine a participant's level of attention to the task. Listeners were instructed to categorize each experimental stimulus (tone pair) as either 'rising' or 'falling', depending on their perception of the sequence. Based on their answers, we calculated their individual 'Coefficient of Sound Perception Preference' (D_p) using the formula $D_p = (F-Sp)/(F+Sp)$, where F is the number of virtual fundamental classifications and Sp the number of spectral classifications. We calculated the 'Listener Attention Coefficient' (D_A) as the proportion of correctly categorized unambiguous stimuli. In order to test the validity of the perceptual test, we repeated the measurement approximately one month later under the same conditions with a subset of the participant set (N=64). The Shapiro-Wilks test of normality revealed that the coefficient D_p was not normally distributed in our experimental group: the majority of our participants performed as fundamental listeners (Mean_D_p=.397, SD=.406). A comparison of the first and the second measurement showed that even without feedback, repeated exposure to the ambiguous stimuli results in a shift towards fundamental auditory bias (Spearman's rho = .69, p < .0001). In order to explore the possibility that the difference between the first and the second measurement of participants' sound perception preference was due to the level of attention devoted to the task, we compared the absolute difference between the first and the second D_p to the attention coefficient D_A. The correlation between the attention coefficient D_A in the first measurement and the $| D_{p1} - D_{p2} |$ was significant (Spearman's rho = -.35, p < .01), indicating that poor attention to the task in general may have been the reason for the observed shift in D_p. In the subsequent analysis relating speakers' perceptual bias to their ability to imitate F0 in a shadowing task, we used the value of D_p collected during the first measurement (i.e., in the same session as the shadowing task). We excluded the data from participants whose attention coefficient D_A was lower than .9, i.e. those who classified two or more of the unambiguous stimuli incorrectly; this results in a sample of N=41.

Shadowing task. The material used during the shadowing task consisted of 16 sentences (8 declaratives and 8 interrogatives) that were presented four times in different orders to each participant. During the first and the fourth presentation, the sentences were presented one-by-one on a computer screen and the participant was instructed to read them in a neutral manner. During the second and the third presentation, the material was presented in an auditory form through a Sennheiser HMD26-600 headset. The participant was asked to repeat the sentences as precisely as possible. For model talkers, we used four different Dutch speakers (two male and two female) who were selected from a set of ten candidates on the grounds of speech clarity and lack of regional accent. The model talkers pre-recorded the 16 sentences in a soundproof booth with a Sennheiser HMD26-600 mic headset. Per participant, we used the recordings of a single model talker in order to increase exposure to the model talker's pitch. The participants were randomly divided between two experimental conditions; half of the participants heard full speech recordings while the other half heard recordings that were filtered with a 300-Hz high-pass Butterworth filter implemented in the Signal Processing Toolbox in Matlab. We calculated the 'Degree of F0 Imitation' by subtracting the absolute difference between the second and third F0 measurement (where the participant was shadowing) from the absolute difference between the model talker's F0 and the participant's F0 in the first measurement (baseline). Thus, a positive value of the 'Degree of F0 Imitation' indicates that the participant adapted to the model talker's F0, a negative value means the participant diverged and 0 represents no measurable change in mean F0. Given the considerable size of the speech corpus that was collected (5632 utterances in total), for the purpose of the analysis reported here, we focused solely on the initial voiced segment in one of the interrogative sentences ('Kan ik bij een vertraging mijn geld terug krijgen?' - "Can I get my money back in case of delays?"). The segment was extracted using a semi-automatic method and the pitch was determined with the autocorrelation method in Matlab.

The F0 measurements were analyzed with multiple regression with the between-subject condition F0 Filter (full speech signal vs. signal with frequencies above 300 Hz) and the participant's Coefficient D_p as predictors and the Degree of F0 Imitation as the dependent variable. As shown in Table 1 and 2, respectively, there was a main and interaction effect of the two predictors on the Degree of F0 Imitation, both for the second measurement (the first time the participants were shadowing the model talker's speech), F(2,38)=3.443, p<.05, as well as for the

Table 1. Summary of Multiple Regression Analysis for the Degree of F0 Imitation in the second measurement (N=41).

Variable	B	SE(B)	Beta	t	Sig.(p)
F0 Filter	50.71	16.20	.591	3.130	.003
Coefficient D_p	51.36	18.31	.443	2.805	.008
F0 Filter * Coefficient ∂_p	-81.08	36.62	-.400	-2.214	.033

Note. r^2=0.252

Table 2. Summary of Multiple Regression Analysis for the Degree of F0 Imitation in the third measurement (N=41).

Variable	B	SE(B)	Beta	t	Sig.(p)
F0 Filter	55.85	19.55	.558	2.857	.007
Coefficient D_p	43.96	22.10	.325	1.989	.054
F0 Filter * Coefficient ∂_p	-61.10	44.19	-.259	-1.383	.175

Note. r^2=0.199

third measurement (the second shadowing), $F(3,37)=3.065$, $p < .05$. This result shows that more fundamental listeners are better in F0 imitation than less fundamental (spectral) listeners, especially in conditions where the F0 information is missing and needs to be derived from the speech signal.

Concluding Discussion

In the study reported here, we used the standard psychoacoustic perceptual test with missing fundamental frequencies to determine a speaker's listener bias. In general, a listener's ability to perceive the missing F0 plays an important role in sound perception in that it helps to track prosodic contours in speech even when they are masked by noise or not transmitted [19], as in phone speech where the region up to 300 Hz is missing. The outcome of the task is somewhat dependent on the composition of the stimuli with the missing fundamental (the number of harmonics favors either the fundamental - for 4 harmonics, or the spectral - for 2 harmonics - interpretation; see [20]). Given that of the 18 ambiguous trials, 12 consisted of 3 and 4 harmonics, the task as such may have favored fundamental perception bias, which might explain the skewed distribution of the coefficient ∂_p in our participant group. A future study may therefore involve a more balanced composition of the stimuli, possibly also varying their length [21]. A comparison of the coefficient ∂_p measured at two different moments within a one-month period shows that the coefficient is not stable. In particular, we observed a slight shift towards a more pronounced fundamental perception, which correlated with the participant's attention to the task. Interestingly, a similar training-independent increase in the salience of the virtual fundamental pitch has been earlier reported by [22], who ascribed it to learning-induced long-term plasticity reflecting the biological relevance of pitch sensation [23]. We subsequently collected speech data in the classical shadowing task with two conditions, one with a full speech signal and one with high-pass filtered speech above 300 Hz. In both conditions, speakers with fundamental listener bias adapted more to the F0 of the model talker; as might be expected, the effect was more pronounced in the high-pass filtered condition. This result suggests advantages for fundamental listeners in communicative situations where F0 imitation is used as a behavioral cue. Future research needs to determine to what extent auditory perception bias may be influenced by training and whether it affects other social processes that rely on parsing of the prosodic information.

References

1. Juslin, P.N., Laukka, P. (2003). Communication of emotions in vocal expression and musical performance: different channels, same code? *Psychological Bul.*, **129**, 770-814.
2. Ververidis, D., Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Com.* **48**, 1162-1181.
3. Gregory, S.W. Jr. (1983). A quantitative analysis of temporal symmetry in microsocial relations. *American Sociological Review* **48**, 129-135.
4. Gregory, S.W. Jr., Hoyt, B.R. (1982). Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *J. Psycholing. Res.* **11**, 35-46.

5. Gregory, S.W. Jr., Webster, S.W., Huang, G. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language and Communication* **13**, 195-217.

6. Gregory, S.W. Jr., Webster, S.W. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *JPSP* **70**, 1231-1240.

7. Gregory, S.W. Jr., Dagan, K., Webster, S.W. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *J. Nonverb. Beh*. **21**, 23-43.

8. Gregory, S.W. Jr., Gallagher, T.J. (2002). Spectral analysis of candidates' nonverbal vocal communication: Predicting U.S. presidential election outcomes. *Social Psychological Quarterly* **65**, 298-308.

9. Babel, M. (2009). *Phonetic and social selectivity in speech accommodation.* Ph.D. diss., University of California, Berkeley.

10. Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review* **105**, 251-279.

11. Babel, M., Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, 1-17.

12. Rousseau, L., Peretz, I., Liégeois-Chauvel, C., Demany, L., Semal, C., Larue, S. (1996). Spectral and virtual pitch perception of complex tones: An opposite hemispheric lateralization? *Brain and Cognition* **30**, 303-308.

13. Schneider, P., Wengenroth, M. (2009). The neural basis of individual holistic and spectral sound perception. *Contemporary Music Review* **28**, 315-328.

14. Wong, P.C.M., Perrachione, T.K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics* **28**, 565-585.

15. Wong, P.C.M., Warrier, C.M., Penhune, V.B., Roy, A.K., Sadehh, A., Parrish, T.B., Zatorre, R.J. (2008). Volume of left Heschl's gyrus and linguistic pitch learning. *Cerebral Cortex* **18**, 828-836.

16. Smoorenburg, G.F. (1970). Pitch perception of two-frequency stimuli. *J. Acoust. Soc. Am.* **48**, 924-942.

17. Laguitton, V., Demany, L., Semal, C., Liégeois-Chauvel, C. (1998). Pitch perception: A difference between right- and left-handed listeners. *Neuropsychologia* **36**, 201-207.

18. Schneider, P., Sluming, V., Roberts, N., Bleeck, S., Rupp, A. (2005). Structural, functional and perceptual differences in Heschl's gyrus and musical instrument preference. *Ann. N.Y. Acad. Sci*. **1060**, 387-394.

19. Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R.D., Seither, S., Lütkenhöner, B. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and non-musicians. *J. Exp. Psychol.: Hum. Percept. Perform*. **33**, 743-751.

20. Houtsma, A.J.M. (1979). Musical pitch of two-tone complexes and predictions by modern pitch theories. J. *Acoust. Soc. Am*. **66**, 87-99.

21. Beerends, J.G. (1989). *Pitches of simultaneous complex tones*. Unpublished Ph.D. diss, University of Eindhoven.

22. Seither-Preisler, A., Johnson, L., Preisler, E., Seither, S., Lütkenhöner, B. (2009). The perception of dual-aspect tone sequences changes with stimulus exposure. *Brain Research Journal* **2**, 125-148.

23. Schneider, P., Sluming, V. Roberts, N., Scherg, M., Goebel, R., Specht, H.J., Dosch, H.G., Bleeck, S., Stippich, C., Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience* **8**, 1241-1247.