Auditory perception bias in speech imitation

Marie Postma-Nilsenová* and Eric Postma

In an experimental study, we explored the role of auditory perception bias in vocal pitch imitation. Psychoacoustic tasks involving a missing fundamental indicate that some listeners are attuned to the relationship between all the higher harmonics present in the signal, which supports their perception of the fundamental frequency (the primary acoustic correlate of pitch). Other listeners focus on the lowest harmonic constituents of the complex sound signal which may hamper the perception of the fundamental. These two listener types are referred to as fundamental and spectral listeners, respectively. We hypothesized that the individual differences in speakers’ capacity to imitate \( F_0 \) found in earlier studies, may at least partly be due to the capacity to extract information about \( F_0 \) from the speech signal. Participants’ auditory perception bias was determined with a standard missing fundamental perceptual test. Subsequently, speech data were collected in a shadowing task with two conditions, one with a full speech signal and one with high-pass filtered speech above 300 Hz. The results showed that perception bias toward fundamental frequency was related to the degree of \( F_0 \) imitation. The effect was stronger in the condition with high-pass filtered speech. The experimental outcomes suggest advantages for fundamental listeners in communicative situations where \( F_0 \) imitation is used as a behavioral cue. Future research needs to determine to what extent auditory perception bias may be related to other individual properties known to improve imitation, such as phonetic talent.

Keywords: pitch, fundamental frequency, imitation, Heschl’s gyrus, missing fundamental

INTRODUCTION

Due to a plethora of linguistic and social functions, vocal pitch imitation plays a central role in human interaction. In language use, pitch, the perceptual correlate of fundamental frequency (\( F_0 \)) typically located between 50–500 Hz in human speech signal, encodes linguistic information regarding speech act and sentence types (Nilsenová, 2007), information structure, and, in many languages, lexical meanings (Ladd, 1996). Pitch imitation arguably accelerates acquisition of these linguistic functions because it is faster than an individual, i.e., trial-and-error based, discovery (Meltzoff et al., 2009). Imitation of phonetic features has also been found to improve speech comprehension (Adank et al., 2010). Listeners who mimicked a novel pronunciation of a sentence improved their subsequent speech reception thresholds for the sentence in a condition with background noise. Next to its linguistic functions, pitch is also the most important vocal source of information regarding emotions, stands and attitudes of the speaker (Juslin and Laukka, 2003; Ververidis and Kotropoulos, 2010). Listeners who mimicked a novel pronunciation of a sentence in a condition with background noise. Next to its linguistic functions, pitch is also the most important vocal source of information regarding emotions, stands and attitudes of the speaker (Juslin and Laukka, 2003; Ververidis and Kotropoulos, 2010). The \( F_0 \) region provides acoustic information for imitation exploited in promoting social convergence and status accommodation (Gregory and Hoyt, 1982; Gregory, 1983; Gregory et al., 1993; Gregory and Webster, 1996; Gregory et al., 1997; Haas and Gregory, 2005; Pardo, 2006) and expressing ingroup–outgroup bias (Babel, 2009; Pardo et al., 2012). Speakers who are perceived as attractive, likable and/or dominant influence listeners’ pitch output, and pitch convergence can be seen as an indicator of cooperative behavior in communication dyads (Nilsenová and Swerts, 2012; Okada et al., 2012). Pitch divergence, on the other hand, suggests that speakers may wish to be viewed as dissimilar and increase social distance between themselves (Giles, 1973). The capacity to perceive the fundamental frequency in the speech signal correctly and to adapt one’s own pitch production according to one’s linguistic and social goals is thus a core communicative skill (Giles and Coupland, 1991).

The results of a range of experimental studies suggest that speakers effortlessly imitate and converge to the phonetic properties of recently heard speech (Natale, 1975; Shockley et al., 2004; Pardo, 2006; Delvaux and Soquet, 2007; Gentilucci and Bernardis, 2007; Nielsen, 2011), including pitch (Goldinger, 1998; Babel and Bulatov, 2012; Gorisch et al., 2012). However, as noted by Babel and Bulatov (2012), in the context of the standard shadowing paradigm, large individual differences can be found in the degree of pitch imitation—with only some participants actually converging to the \( F_0 \) of the model talker (Babel and Bulatov, 2012, p. 240). The proposal of our study is that individual variation in the imitation of pitch is, at least partly, due to basic acoustic perceptual mechanisms that also influence pitch production.

Most speech imitation studies assume that there exist few individual differences among healthy hearing subjects with respect to the low-level processing of speech signal. However, past psychoacoustic research involving stimuli with a missing fundamental indicated that there is a difference between two auditory perceptual extremes, sometimes referred to as analytic and holistic/synthetic listeners (von Helmholtz, 1885; Smoorenburg, 1970;
Houtsma, 1979), henceforth referred to as spectral and fundamental listeners, respectively. Spectral listeners primarily focus on the individual harmonic constituents, they “decompose the sound” (Schneider and Wengenroth, 2009, p. 316), while fundamental listeners are attuned to the relationship between all the higher harmonics present in the signal, which supports their perception of the fundamental frequency (Rousseau et al., 1996; Laguittion et al., 1998; Seither-Preisler et al., 2007). According to von Helmholtz (1885), for fundamental listeners, it is as if the harmonics “fuse into the whole mass of musical sound” (Schneider and Wengenroth, 2009), hence his choice of the term “holistic” or “synthetic” to refer to this type of listening mode. While in practice, few listeners perform uniquely at the absolutes of one or the other type (Ladd et al., 2013), the perceptual bias may lead to different interpretations of perceived pitch values in particular contexts. On the one hand, the perception of the fundamental frequency is supported by so-called combination tones generated in the cochlea (Plomp, 1976). These tones differ across individual listeners (Probst et al., 1986). On the other hand, results of structural MRI studies suggest that the bias is, at least partly, due to a right-/leftward asymmetry of gray matter volume in the lateral Heschl’s gyrus (Schneider et al., 2005a,b; Wong et al., 2008), the so-called “pitch processing center” (Griffiths, 2003). In particular, larger volumes of right Heschl’s gyrus seem to be associated with spectral perceptual bias, while the left Heschl’s gyrus has been linked to changes in the F0 modulation and temporal information (Schneider et al., 2005a,b; Warrier et al., 2009). Until fairly recently, the perceptual bias has mainly been examined in the context of musical psychoacoustics. The research outcomes of Wong et al. (2008), however, may be interpreted as support for the claim that it may also affect linguistic performance. In their study of lexical tone perception, listeners who performed worse in a word identification task involving vowels with superimposed tones showed a smaller Heschl’s gyrus volume on the left than listeners who performed better. Given the tight link between perception and production, recently implemented in the “forward-model” of Pickering and Garrod (2013) where internal simulation of input utterances facilitates comprehension and shapes phonetic output, we assume that advantages in the perception of F0 might improve its imitation. In other words, fundamental listeners may have a better capacity to adapt their pitch to their communication partners than spectral listeners.

In what follows, we present the results of a production study conducted to determine the effect of auditory perception bias on automatic pitch imitation in a classical shadowing task. Listeners’ perception bias was determined with the help of missing fundamental stimuli, an idea that originated with Smoorenburg (1970) who introduced a forced-choice task involving sequences of two complex tones. In the task, participants are presented with a sequence and asked to indicate if the perceived pitch is rising or falling. The crux of the task is that the tone sequence is designed to have an ambiguous pitch change. Each complex tone is created from m partials $F_n, F_{n+1}, \ldots, F_{n+m-1}$, (n is an integer, $n > 0$), without the fundamental $F_0$. The ambiguity arises from the opposite changes of the (missing) fundamentals ($F_0$) and the (physically present) lowest partials ($F_{lp}$). When the subsequent fundamentals $F_0$ are rising, the lowest partials $F_{lp}$ are falling, and vice versa. Representing the partials of the first and second tones by $F^1$ and $F^2$, respectively, fundamental listeners will perceive the change in pitch $\Delta P_f$ by computing $\Delta P_f = (F^1_{lp} + \Delta F^1_{lp}) - (F^2_{lp} - \Delta F^2_{lp})$ (k $\in \{n, n+1, \ldots, n+m-2\}$) in order to estimate $F^2_{lp} - F^1_{lp}$. Spectral listeners will rely on $\Delta P_{lp} = F^2_{lp} - F^1_{lp}$ to determine if the pitch is rising or falling. Figure 1 illustrates an ambiguous tone sequence. The sequence depicted has a falling $F_0$ ($\Delta P_f < 0$) and a rising $F_{lp}$ ($\Delta P_{lp} > 0$).

Since the early work of Smoorenburg (1970), the listener type task has been frequently employed to study how acoustic variables (e.g., $F_0$-value, $\Delta P$-value, number of partials) affect the perception of pitch (Plomp, 1965; Ladd et al., 2013). Schneider et al. (2005b) and Seither-Preisler et al. (2007) used the task to explore the distribution of listener types in relation to musical training. In both studies, participants were presented with a number of ambiguous-sequence stimuli. The proportion of stimuli to which a fundamental or spectral pitch change was perceived by the participants, defined the so-called Coefficient of Sound Perception Preference ($\delta_P$), a value ranging from $-1$ (all stimuli perceived as spectral) to $+1$ (all stimuli perceived as fundamental). To prevent the emergence of combination tones that arise at the level of the cochlea (Terhard, 1974), Schneider et al. (2005b) presented tones at low intensity and Seither-Preisler et al. (2007) added masking noise to their stimulus sequences. Given that the perception thresholds of combination tones vary with the individual (Plomp, 1976), following Ladd et al. (2013), we made use of stimuli without masking in an attempt to include possible effects of cochlear mechanisms on the perception and production of pitch in speech.

**MATERIALS AND METHODS**

**PARTICIPANTS**

Eighty-eight Dutch native speakers (67 females) between the age of 17–25 years ($M = 20.48, SD = 2.12$) participated in the experiment for course credit. None of them reported any hearing

---

**FIGURE 1 | Illustration of an ambiguous two-tone sequence to determine auditory perception bias.** The sequence has a falling (missing) fundamental $F_0$ and a rising lowest partial $F_{lp}$. The horizontal lines represent the partials of the tone. The higher partials are physically present (solid lines) and the lower partials are missing (dashed lines).
difficulties. Fourteen of the participants were left-handed; about one half of the experimental group described their musical proficiency as low to average, the other half assessed their proficiency as high to professional. Male and female were divided equally between the two experimental conditions. Prior to the experiment which had received an approval from the ethical committee, participants provided their written informed consent.

MEASURING AUDITORY PERCEPTION BIAS
Participants’ auditory perception bias was determined with a variation of the psychoacoustic perceptual test described in Smoorenburg (1970), Laguitton et al. (1998), Schneider et al. (2005b), and Seither-Preisler et al. (2007). 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 Laguitton et al. (1998). Participants were asked to categorize 18 perceptually ambiguous stimuli sequences consisting of two complex tones, tone 1 and tone 2 as illustrated in Figure 1. All tones 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). The other 18 stimuli served as control trials in that their interpretation was unambiguous but helped 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” ($\delta_p$) using the equation $\delta_p = (F - Sp)/(F + Sp)$, where $F$ is the number of virtual fundamental classifications and $Sp$ the number of spectral classifications. We calculated the “Listener Attention Coefficient” ($\delta_A$) as the proportion of correctly categorized ambiguous stimuli. In order to test the validity of the perceptual test, we repeated the measurement approximately 1 month later under the same conditions with a subset of the participant set ($N = 64$). In the analyzes presented below, we report the overall results for all experimental stimuli ($\delta_p$), as well as the results for stimuli where the lowest present component frequency $F_n > 1000$ Hz, $\delta_{1000}$. The 1000 Hz value is arguably the highest frequency at which $F_0$ could be produced by a human voice and also the approximate maximal value at which the missing fundamental phenomenon occurs (Fletcher, 1924). Stimuli with $F_n > 1000$ Hz thus arguably support the perception of the missing fundamental.

SPEECH IMITATION TASK
The shadowing task took place immediately after the psychoacoustic task. It consisted out of eight declarative and eight interrogative sentences uttered by four different model talkers (two male, two female) in a between-subject design in order to maximize exposure to the model speaker’s voice and thus increase chances of possible imitation. The 16 sentences were recorded four times: in the first and fourth block, the participants read the sentences in a randomized order (same for all participants) from a PowerPoint slide; the declarative and interrogative sentences were presented in a mixed design. In the second and third block, they were asked to repeat the sentences as they were presented to them (in auditory modus only), through high quality headphones (Sennheiser HMD26–600–7). The participants were not explicitly instructed to imitate the speakers’ pronunciation but simply to repeat the utterances. They were randomly assigned to one of two between-subject conditions (filtered vs. unfiltered). In the filtered condition, participants heard recordings that were filtered with an order nine high-pass Butterworth filter with cutoff frequency of 300 Hz, using Matlab’s Signal Processing Toolbox. The participants in the unfiltered condition heard full speech recordings.

AUTOMATIC PITCH ESTIMATION
An initial set of analyzes was performed on the whole corpus with a subsequent more detailed analysis of a shorter speech segment. The recordings were segmented per utterance and analyzed using the autocorrelation method, see, e.g., Rabiner et al. (1976), implemented in Matlab using a frame length of 10 ms with 5 ms overlap, and a frequency range of 50–500 Hz. For the whole corpus, we computed five statistical descriptors of $F_0$: the mean value, the maximum, the minimum, the range (max–min) and the standard deviation. The degree of $F_0$ imitation was determined by assessing the z-score of the absolute difference between the model speaker’s $F_0$ descriptor and the participant’s $F_0$ descriptor in the first block ($D_1$, baseline) and the second and third block (first and second shadowing, $D_2$ and $D_3$, respectively). We defined two measures of imitation, $F_0$ Imitation1 = $D_1 - D_2$ and $F_0$ Imitation2 = $D_1 - D_3$. The statistical analyses were conducted with the IBM SPSS Statistics software v.2.0.

RESULTS
In this section we present the results of our experiment in four parts. First, we present the descriptive values of the “Coefficient of Sound Perception Preference” $\delta_p$ in the first and second measurement. Second, all results obtained in the first measurement are compared to global—sentence level—imitative behavior. Third, using smaller speech segments, a correlation analysis is performed on the psychoacoustic and socio-demographic variables to determine the inclusion of variables in a regression analysis. Finally, the results of a hierarchical multiple regression analysis are presented that relate auditory perception bias to $F_0$ imitation.

COEFFICIENT OF SOUND PERCEPTION PREFERENCE
The Shapiro–Wilks test of normality revealed that the coefficient $\delta_p$ was not normally distributed: the majority of the participants performed as fundamental listeners (Mean $\delta_p = 0.397$, SD = 0.406). For a distribution of the $\delta_p$, see Figure 2. A comparison of the first and the second measurement showed that repeated exposure to the ambiguous stimuli resulted in a shift toward the fundamental bias, with a significant correlation between the two measurements (Spearman’s $\rho = 0.69$, $p < 0.001$). The test-retest correlation was comparable to that provided by Ladd et al. (2013). The difference between the two measurements was marginally significant with Wilcoxon Signed Ranks test, $Z = -1.87$, $p = 0.06$. In order to explore the possibility that the difference between the first and the second measurement of participants’ perception was due to the level of attention devoted to the task, we compared the absolute difference between the first
and the second $\delta_p$, $\delta_p1$, and $\delta_p2$, to the attention coefficient $\delta_A$. The correlation between the attention coefficient $\delta_A$ in the first measurement and the $|\delta_p1 - \delta_p2|$ was significant (Spearman’s $\rho = -0.35$, $p < 0.01$), indicating that poor attention to the task during the first measurement may have been the reason for the observed shift in $\delta_p$ (given that the shift was in the direction from “undecided” to a more “pure” type of perception, see Figure 2).

As pointed out by Seither-Preisler et al. (2007), however, who reported a similar result attributed to repeated exposure, an effect due to learning cannot be excluded (no measures of $\delta_A$ were provided in their study). In the subsequent analysis relating speakers’ perceptual bias to their capacity to imitate $F_0$, we used the value of $\delta_p$ collected during the first measurement, i.e., in the same session as the shadowing task.

**SENTENCE-LEVEL IMITATION**

In the initial global analysis of the whole corpus, for each pitch value, we conducted two statistical tests, one with the full participant sample and one where we only included those participants who had more than 90% correct (less than 2 mistakes of the total of 18 trials) in the categorization of the unambiguous stimuli in the psychoacoustic task ($N = 41$ in total, with $N_{\text{full signal}} = 22$ and $N_{\text{filtered}} = 19$) and were thus assumed to be reliable as listeners. Tables 1, 2 give an overview of the correlations between $F_0$ imitation (rows expressed in Hz) for the five descriptors (columns) and the Coefficient of Sound Perception Preference, $\delta_p$, split by condition (with full signal, viz. Table 1, and with the signal under 300 Hz filtered out, viz. Table 2).

The results of the global analyses suggested that, overwhelmingly, participants who performed reliably on the non-ambiguous task and scored higher in the direction of fundamental listeners imitated the model speakers’ pitch to a higher degree, especially in the condition with filtered speech signal. Given that the analyses were performed on full utterances, however, they might have been less likely to capture $F_0$ imitation that typically occurs on individual segments (especially, vowels) and less
reliable given that local minima and maxima (that, in turn, affect the range and SD) may be outliers in the signal without communicative significance. Therefore, we proceeded with a more fine-grained analysis of a subset of the corpus, in which we also included socio-demographic variables collected in the experiment.

Vowel Segment Imitation

In order to limit the size of the corpus collected in the shadowing task, we randomly selected one of the interrogative sentences for the subsequent analyses, focusing on its initial voiced segment. The choice of an interrogative sentence was driven by the assumption that imitation is likely to occur at sentence-initial boundaries immediately following the model talker’s output (Nilsenova and Nolting, 2010), and (2) polar (yes/no-) interrogatives that are context-free (no particular word in the interrogative is in focus) are intonation-marked by a pitch excursion (van Heuven and Haan, 2002), i.e., in this case, on the finite verb that is sentence-initial due to subject-verb inversion. An automatic analysis of pitch was performed on the initial occurrence of the vowel /a/ in the sentence. The segment fundamental frequency was determined by averaging over the $F_0$ values of approximately the first half of the initial vowel in order to avoid right vowel boundary detection errors.

Preliminary data analysis was conducted to identify potential covariates, using both demographic and psychoacoustic variables. Chi-square tests indicated that there were no significant differences between the full speech and high-pass filtered condition with respect to participant gender and handedness, there was also no significant difference between stimulus voice (two male, two female) and participant gender. Non-parametric Mann–Whitney Tests for variables without normal distribution indicated no significant difference between the experimental conditions with respect to musicality [determined on the basis of a self-reported session) in the two experimental conditions.]

Table 3 | Zero-order Pearson product-moment correlations among psychoacoustic variables and the socio-demographic variables.

| Variable                | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|-------------------------|----|----|----|----|----|----|----|----|----|----|
| 1. Condition            | −  |    |    |    |    |    |    |    |    |    |
| 2. Age                  | −0.02 | −  |    |    |    |    |    |    |    |    |
| 3. Gender               | 0.08 | −0.14 | −   |    |    |    |    |    |    |    |
| 4. Handedness           | −0.06 | 0.01 | 0.12 | −  |    |    |    |    |    |    |
| 5. Musicality           | −0.07 | 0.18 | 0.07 | −0.18 | −  |    |    |    |    |    |
| 6. $\delta_A$           | 0.06 | 0.07 | −0.07 | −0.04 | 0.49** | −  |    |    |    |    |
| 7. $\delta_p$           | −0.09 | 0.11 | 0.04 | 0.09 | 0.51** | 0.47** | −  |    |    |    |
| 8. $\delta_{p1000}$     | −0.10 | 0.06 | 0.01 | 0.04 | 0.46** | 0.39** | 0.94** | −  |    |    |
| 9. F0 Imitation$_1$      | 0.19‡ | −0.03 | 0.11 | 0.12 | −0.03 | 0.05 | 0.17 | 0.24* | −  |    |
| 10. F0 Imitation$_2$     | 0.15‡ | −0.05 | 0.18‡ | 0.10 | 0.00 | 0.06 | 0.17 | 0.18 | 0.62** | −  |

*Condition* was dummy-coded to compare the effect of frequency filtering with other responses (1 = filtered, 0 = full speech). *Gender* was dummy-coded to compare the performance of male and female listeners (1 = female, 0 = male). *Handedness* was dummy-coded to compare the performance of left- and right-handers (1 = right, 0 = left). $\delta_A$, Listener Attention Coefficient; $\delta_p$, Coefficient of Sound Perception Preference; $\delta_{p1000}$, Coefficient of Sound Perception Preference above 1000 Hz.

$^a_{p < 0.10, \ *p < 0.05, \ **p < 0.001}$.
Hierarchical multiple regression was used to establish the incremental value of auditory perception bias when predicting the level of $F_0$ imitation in a condition with high-pass band filtered speech and in a condition with full speech signal. The regression model consisted of two blocks and assessed the additional variance explained with the estimation of each added block. At Block 1, the centered values of $\delta_{p1000}$ and experimental condition were entered simultaneously. This block resulted in a significant overall model, $F(2, 85) = 4.87, p = 0.01$, accounting for 10% of the variance in the imitation scores. The interaction effect between $\delta_{p1000}$ and experimental condition was created by multiplying the mean-centered values of each individual variable and then was entered at Block 2 along with all variables entered at Block 1. Results again indicated an overall effect for the model, $F(3, 84) = 3.27, p = 0.03$, explaining an additional variance of 0.2%. The $\delta_{p1000}$ by experimental condition interaction term did not significantly predict the imitation scores after controlling for covariates and main effects ($b = -8.02, p = 0.69$). Figures 3, 4 graphically display the main effects of $\delta_{p1000}$ and condition on $F_0$ imitation. The y-axes express the difference between $D_1$, the absolute difference between the model speaker’s $F_0$ and the participant’s $F_0$ in the first (baseline) block, and $D_2$, the absolute difference between the model speaker’s $F_0$ and the participant’s $F_0$ in the second (first shadowing) block; a positive value here indicates imitation and a negative value indicate divergence. The figures show that more fundamental listeners were better at imitating the fundamental frequency in the model speaker’s voice. Fully tabulated results of the hierarchical regression model are presented in Table 4.

**DISCUSSION**

Our findings suggest that auditory perception bias can partly account for the individual variation found in earlier pitch imitation studies. In a shadowing task, fundamental listeners showed a better capacity to imitate the vocal pitch of the model talkers, especially in a condition where the region between 0–300 Hz has been filtered out and information about $F_0$ had to be derived from the higher frequencies (akin to telephone speech). These results can be used in future studies on speech imitation abilities, e.g., to explore phenomena such as phonetic (pronunciation) talent (Lewandowski, 2009).

Our findings of individual differences in listener’s sensitivity to tone sequences may be related to those of Semal and Demany (2006), who found some listeners to be able to detect changes in tone sequences, while unable to indicate the direction of change (upward or downward). Future studies should address the relation between individual differences in sensitivity to pitch direction and in auditory perception bias. At this point it is unclear what is causing the individual differences in

**TABLE 4 | Results of the hierarchical regression model.**

| Variable | $b$  | $SE$ | $\beta$ | Adjusted $R^2$ | $\Delta R^2$ |
|----------|------|------|---------|----------------|-------------|
| Step 1   |      |      |         |                |             |
| $\delta_{p1000}$ | 26.07 | 10.02 | 0.26** | 0.08           | 0.10**      |
| Filter condition | 17.46 | 8.33  | 0.22*   |                |             |
| Step 2   |      |      |         |                |             |
| $\delta_{p1000}$ | 24.79 | 10.10 | 0.26*   | 0.07           | 0.00*       |
| Filter condition | 17.44 | 8.37  | 0.22*   |                |             |
| $\delta_{p1000}$ by filter condition | -8.02 | 20.19 | -0.04 |                |             |

$\delta_{p1000}$, coefficient of sound perception preference above 1000 Hz.

*p $\leq 0.05$, **p $\leq 0.01$. 

[Image 312x298 to 539x472]
auditory perception bias. As stated in the Introduction, Schneider et al. (2005b) found neuroanatomical differences in the lateral Heschl’s gyrus to be associated with perception bias. However, the differences may very well be of a more peripheral origin, i.e., reflecting individual differences in cochlear responses. In particular, non-linear interactions in the cochlea may give rise to so-called combination tones (Plomp, 1965). When stimulated with a tone consisting of the $n$-th and $(n + 1)$th harmonic, the cochlea may generate tones at a frequency corresponding to that of the missing fundamental. It is important to stress that the generated tone is physically present because it is generated in the cochlea, rather than being extracted from the harmonics (as is the case for the missing fundamental). Plomp (1965) claimed that combination tones are inaudible for “usual levels” of speech and music and that the same applies to the perception of the missing fundamental. Notwithstanding this claim, in his study of individual differences in (what we call) auditory perception bias, Smoorenburg (1970) effectively suppressed the perception of combination tones by superimposing masking noise bands centered at the combination-tone frequencies. Apparently, Smoorenburg (1970) was concerned about a potential interfering effect of combination tones in the determination of listener type. Given that in the experiment reported here, the stimuli were presented without masking noise, the participants may have perceived physically generated tones at the level of the missing fundamental. The generation of combination tones could have lead to overestimates of $\delta_p$, because spectral listeners may perceive the combination tone instead of a reconstructed fundamental (as fundamental listeners do), thus explaining the skewed distribution in both first and second measurement of the perception bias. On the one hand, the presence of combination tones may invalidate the determination of listener type. On the other hand, combination tones are an inevitable byproduct of naturally occurring sounds. Cochlear dynamics generate combination tones which affect further cortical processing and anatomical correlates (i.e., lateral Heschl’s gyrus). As such, the auditory perception bias as measured in our experiment takes into account individual variations in sensitivity to combination tones. In general, the potential role of combination tones in the definition and study of listener types deserves further attention. Ladd et al. (2013) pointed at the methodological differences in earlier studies of listener type performed by Schneider et al. (2005b) and Seither-Preisler et al. (2007), but did not identify the use of masking noise (or other means to suppress combination tones) as a main methodological difference between their study and both earlier ones. In our future work, we aim at a detailed investigation of the role of combination tones in auditory perception bias.

ACKNOWLEDGMENTS

The authors would like to thank Bob Ladd, Annemarie Seither-Preisler, and Peter Schneider for useful discussions. Jeffrey Vollenbregt is acknowledged for his help with collecting the experimental data and Shanook Janssen for her help with segmentation of the recordings.

REFERENCES

Adank, P., Hagoort, P., and Bekkering, H. (2010). Imitation improves language comprehension. Psychol. Sci. 21, 1903–1909. doi: 10.1177/0956797610389192

Bavel, M. (2009). Phonetische und Sozial Selektivität in Speech Accommodation. PhD thesis, University of California.

Bavel, M., and Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. Lang. Speech 55, 231–248. doi: 10.1177/0028380911417695

Delvaux, V., and Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. Phonetica 64, 145–173. doi: 10.11159/001017914

Fletcher, H. (1924). The physical criterion for determining the pitch of a tone. Phys. Rev. 23, 427–437. doi: 10.1103/PhysRev.23.427

Gentilucci, M., and Bernardis, P. (2007). Imitation during phoneme production. Neuropsychologia 45, 608–615. doi: 10.1016/j.neuropsychologia.2006.04.004

Giles, H. (1973). Accent mobility: a model and some data. Anthropol. Linguist. 15, 87–105.

Giles, H., and Coupland, N. (1991). Language: Contexts and Consequences. Milton Keynes: Open University Press.

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. Psychol. Rev. 105, 251–279. doi: 10.1037/0033-295X.105.2.251

Gorisch, J., Wells, B., and Brown, G. (2012). Pitch contour matching and interaction alignment across turns: an acoustic investigation. Lang. Speech 55, 57–76. doi: 10.1177/0028380911428874

Gregory, S. (1983). A quantitative analysis of temporal symmetry in microsocial relations. Am. Soc. Rev. 48, 129–135. doi: 10.2307/2095151

Gregory, S., Dagan, K., and Webster, S. (1997). Evaluating the relation of vocal accommodation in conversation partners’ fundamental frequencies to perceptions of communication quality. J. Nonverbal Behav. 21, 23–43. doi: 10.1023/A:1024995717773

Gregory, S., and Hoyt, B. (1982). Conversation partner mutual adaptation as demonstrated by fourier series analysis. J. Psycholinguist. Res. 11, 35–46. doi: 10.1007/BF01067500

Gregory, S., and Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. J. Personal. Soc. Psychol. 70, 1231–1240. doi: 10.1037/0022-3514.70.6.1231

Gregory, S., Webster, S., and Huang, C. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. Lang. Commun. 13, 195–217. doi: 10.1016/0721-5309(93)90026-J

Griffiths, T. (2003). Functional imaging of pitch analysis. Ann. N.Y. Acad. Sci. 999, 40–49. doi: 10.1196/annals.1284.004

Haas, A., and Gregory, S. (2005). The impact of physical attractiveness on women’s social status and interactional power. Sociol. Forum 20, 449–471. doi: 10.1177/1053455X05280539

Houtsma, A. (1979). Musical pitch of two-tone complexes and predictions by Heschl’s gyrus. J. Acoust. Soc. Am. 66, 87–90. doi: 10.1121/1.382943

Judlin, P., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? Psychol. Bull. 129:770. doi: 10.1037/0033-2909.129.5.770

Ladd, D. (1996). Intonational Phonology. Cambridge, MA: Cambridge University Press.

Ladd, D., Turnbull, R., Browne, C., Caldwell-Harris, C., Ganushchak, L., Swoboda, K., et al. (2013). Patterns of individual differences in the perception of missing-fundamental tones. J. Exp. Psychol. Hum. Percept. Perform. 39, 1386–1397. doi: 10.1037/a0031261

Lagoutitton, V., Demany, L., Semal, C., and Liégeois-Chauvel, C. (1998). Pitch perception: a difference between right- and left-handed listeners. Neuropsychologia 36, 201–207. doi: 10.1016/S0028-3932(97)00122-X

Lewandowski, N. (2009). Sociolinguistic factors in language proficiency: phonetic convergence as a signature of pronunciation talent. Lang. Talent Brain Act. 1, 257.

Meltzoff, A., Kuhl, P., Movellan, J., and Sejnowski, T. (2009). Foundations for a new science of learning. Science 325, 284–288. doi: 10.1126/science.1175626

Natale, M. (1975). Social desirability as related to convergence of tempo-ral speech patterns. Percept. Mot. Skills 40, 827–830. doi: 10.2466/pms.1975.40.3.827

Nielsen, K. (2011). Specificity and abstractness of vowel imitation. J. Phon. 39, 132–142. doi: 10.1016/j.wocn.2010.12.007

Nilsenová, M. (2007). “Nuclear rises in update semantics,” in Questions in Dynamic Semantics, eds M. Aloni, A. Butler, and P. Dekker (Amsterdam: Elsevier), 295–314.
Nilsenová, M., and Nolting, P. (2010). Linguistic adaptation in semi-natural dialogues: age comparison. *Text Speech Dialogue Lect. Notes Comput. Sci.* 6231, 531–538. doi: 10.1007/978-3-642-15760-8_6

Nilsenová, M., and Swerts, M. (2012). "Prosodic adaptation in language learning," in Pragmatics and Prosody in English Language Teaching, ed. J. Romero-Trillo (Berlin: Springer), 77–96. doi: 10.1007/978-94-007-3883-6_6

Okada, B., Lachs, L., and Boone, B. (2012). Interpreting tone of voice: musical pitch relationships convey agreement in dyadic conversation. *J. Acoust. Soc. Am.*, 132, EL208–EL214. doi: 10.1121/1.4742316

Pardo, J. (2006). On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am.* 119, 2382–2393. doi: 10.1121/1.2178720

Pardo, J., Gibbons, R., Suppes, A., and Krauss, R. (2012). Phonetic convergence in college roommates. *J. Phon.* 40, 190–197. doi: 10.1016/j.wocn.2011.10.001

Pickering, M., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495

Plomp, R. (1965). Detectability threshold for combination tones. *J. Acoust. Soc. Am.* 47, 1111–1123.

Plomp, R. (1976). Aspects of Tone Sensation. London: Academic Press.

Probst, R., Coats, A., Martin, G., and Lonsbury-Martin, B. (1986). Spontaneous, click-, and toneburst-evoked otoacoustic emissions from normal ears. *Hear. Res.* 21, 261–275. doi: 10.1016/0378-5955(86)90224-8

Rabiner, L., Cheng, M., Rosenberg, A., and McGonegal, C. (1976). A comparative performance study of several pitch detection algorithms. *IEEE Trans. Audio Signal Speech Process.* 24, 399–417. doi: 10.1109/TASSP.1976.1162846

Rousseau, L., Peretz, I., Liégeois-Chauvel, C., Demany, L., Semal, C., and Larue, S. (1996). Spectral and virtual pitch perception of complex tones: an opposite terms.

Schneider, P., Sluming, V., Roberts, N., Bleeck, S., and Rupp, A. (2005a). Structural, functional and perceptual differences in heschl’s gyrus and musical instrument preference. *Ann. N.Y. Acad. Sci.* 1060, 387–394. doi: 10.1196/annals.1360.033

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H., et al. (2005b). Structural and functional asymmetry of lateral heschl’s gyrus reflects pitch perception preference. *Nat. Neurosci.* 8, 1241–1247. doi: 10.1038/nn1530

Schneider, P., and Wengenroth, M. (2009) The neural basis of individual holistic and spectral sound perception. *Contemporary Music Review*, 28, 315–328. doi: 10.1080/0749446090404402

Seither-Preisler, A., Johnson, L., Krumhholz, K., Nobbe, A., Patterson, R., Seither, S., et al. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 743–751. doi: 10.1037/0096-1523.33.3.743

Semal, C., and Demany, L. (2006). Individual differences in the sensitivity to pitch direction. *J. Acoust. Soc. Am.* 120, 3907–3915. doi: 10.1121/1.2357708

Shockley, K., Sabadini, L., and Fowler, C. (2004). Imitation in shadowing words. *Percept. Psychophys.* 66, 422–429. doi: 10.3758/BF03194890

Smooenbug, G. (1970). Pitch perception of two-frequency stimuli. *J. Acoust. Soc. Am.* 48, 924–942. doi: 10.1121/1.1912232

Terhardt, E. (1974). Pitch, consonance and harmony. *J. Acoust. Soc. Am.* 55, 1061–1069. doi: 10.1121/1.1914648

van Heuven, V., and Haan, J. (2002). Temporal distribution of interrogativity markers in dutch: a perceptual study. *Labor. Phonol.*, 7, 61–86.

Ververidis, D., and Kotropoulos, C. (2006). Emotional speech recognition: resources, features, and methods. *Speech Commun.* 48, 1162–1181. doi: 10.1016/j.specom.2006.04.003

von Helmholz, H. (1885). *On the Sensations of Tone. London: Longmans.*

Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., et al. (2009). Relating structure to function: Heschl’s gyrus and acoustic processing. *J. Neurosci.* 29, 61–69. doi: 10.1523/JNEUROSCI.3489-08.2009

Wong, P., Warrier, C., Penhune, V., Roy, A., Sadeh, A., Parrish, T., et al. (2008). Volume of left heschl’s gyrus and linguistic pitch learning. *Cereb. Cortex* 18, 828–836. doi: 10.1093/cercor/bhm115

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 April 2013; accepted: 17 October 2013; published online: 05 November 2013.

Citation: Postma-Nilsenová M and Postma E (2013) Auditory perception bias in speech imitation. *Front. Psychol.* 4:826. doi: 10.3389/fpsyg.2013.00826

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

Copyright © 2013 Postma-Nilsenová and Postma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.