ISCA Archive
http://www.isca-speech.org/archive

4th International Symposium on
Tonal Aspects of Languages (TAL-2014)
Nijmegen, The Netherlands
May 13-16 2014

No effect of language experience on spectral/fundamental listener type distribution: A comparison of Chinese and Dutch

Marie Postma-Nilsenová, Eric Postma & Yu Gu

Tilburg center for Cognition and Communication, Tilburg University, The Netherlands
m.nilsenova@uvt.nl, e.o.postma@uvt.nl, y.gu_1@uvt.nl

Abstract

Individual listeners differ in their capacity to process complex tones. Past studies suggest that the distribution of listener types, though partly genetically determined, might also depend on the linguistic environment. Listeners raised in a tonal language may be more sensitive to perceive $F0$ changes both in music and speech, due to their vocal experience. Using the missing fundamental task, an earlier study by Ladd et al. (2013) found no differences between English and Chinese listeners, but did not take into account the potentially confounding effect of combination tones that arise in the cochlea. This study determines the listener-type distributions of Dutch and Chinese individuals while employing masking noise to suppress the perception of combination tones. The results show no differences in listener-type distribution in groups of Dutch and Chinese listeners. Simulations performed with a tonal and non-tonal memory-based computational model of pitch perception reveal that in both types of languages the fundamental listening mode is dominant.
Index Terms: Tone, listener type, perception

1. Spectral and Fundamental listeners

Psychoacoustic research originating in the 19th century [1] describes differences between two types of listeners, referred to as analytic and synthetic listeners [2, 3], or spectral and fundamental listeners, respectively. Analytic/spectral listeners primarily focus on information contained in the signal spectrum; synthetic/fundamental listeners, on the other hand, focus on changes in the F_0 modulation and temporal information [4, 5, 6]. While in practice, few listeners perform uniquely at the absolutes of one or the other type [7], the perceptual bias may lead to different interpretations of perceived pitch values in particular contexts. Results of functional MRI and magnetoencephalography studies suggest that the core of the bias is a right-/leftward asymmetry of gray matter volume in the lateral Heschl's gyrus [8, 9, 10], the so called 'pitch processing center' [11].

In a ground-breaking study, Dediu and Ladd [12] found a statistical relation between linguistic tone distribution and two genes. The relation could, possibly, be accounted for by the process of genetic predisposition leading to differences in brain structure and function [13], particularly the asymmetry in the 'pitch processing center'. Even though the exact prediction of [12] was not confirmed by a subsequent experimental study [14], there does appear to be a connection between the capability to process lexical tone information and the brain asymmetry associated with different types of listeners. In particular, in the context of lexical tone perception, fundamental listeners perform better in a categorization task involving vowels with superimposed tones than listeners with a spectral bias (smaller Heschl's gyrus volume on the left) [15]. This finding raises the question to what extent the variation in linguistic tone perception could be due to the "linguistic diet" of the listeners, i.e., exposure to particular kind of tonal input in speech.

In our study, we compared the preferred listening modus, as well as the linguistic input of speakers of two different languages: Dutch (arguably, a non-tonal language from a non-tonal family) and Chinese (a tonal language from a tonal family). In order to determine the prevalence of spectral/fundamental listening modus, we employed the missing fundamental task used in earlier studies. Since the early work of Smoorenburg [3], the missing fundamental task has been frequently employed to study how acoustic variables (e.g., F_0-value, ΔP-value, number of partials) affect the perception of pitch [16, 7]. During the task, participants are presented with a number of ambiguous-sequence stimuli. The proportion of stimuli to which a fundamental or spectral pitch change is perceived by the participants, defines the so-called Coefficient of Sound Perception Preference' (δ_p) [8], a value ranging from -1 (all stimuli perceived as fundamental) to $+1$ (all stimuli perceived as spectral). Given that Chinese speakers make use of subtle changes in $F0$ to convey lexical meanings, we assumed that they would be more likely to perform as fundamental listeners than speakers of Dutch.

The difference between speakers of tonal and non-tonal languages has previously been studied by Ladd et al. [7] who reported no significant difference between the two groups they examined. However, their stimulus material was based on tones that have likely included the effect of so-called combination tones. Typically, a part of the missing fundamental task consists in trying to prevent the emergence of nonlinear interactions in the cochlea that give rise to combination tones [17]. 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 and thus support the impression of $F0$ being present in the signal [16]. 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). Given that in the experiment of [7], the stimuli were presented without masking noise, the participants may have perceived physically generated tones at the level of the missing fundamental, i.e., the generation of combination tones could have lead to overestimates of δ_p.

Plomp [16] 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. A possible manner of excluding their effect is, therefore, by presenting the stimuli at low intensity (as in, e.g., [8]). Due to the fact that sensitivity to the intensity is highly individual (as are, incidentally,

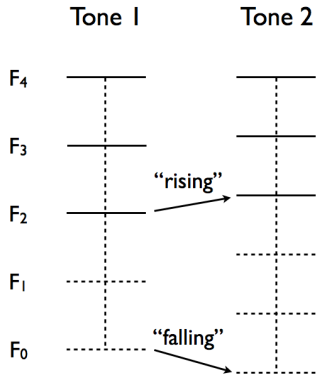

Figure 1: Illustration of an ambiguous two-tone sequence to determine auditory perception bias. The sequence has a falling (missing) fundamental F_0 and a rising lowest partial F_2.

the exact combinations of tones that arise in the cochlea), it may be preferable to exclude the combination tone effect in another way. In fact, in his study of individual differences in missing $F0$ perception, Smoorenburg [3] effectively suppressed the effect of combination tones by superimposing masking noise bands centered at the combination-tone frequencies, a method also employed by [4] and in the current study. In what follows, we first present the results of a perceptual task and, subsequently, report the outcomes of a computational modelling approach.

2. Perceptual Task

2.1. Methods

Participants. In total, 60 Chinese native speakers (26 male, M_{age} = 26.5, SD = 4.4) and 70 Dutch native speakers (28 male, M_{age} = 23.3, SD = 5.5) participated in the study in exchange for course credit or a small gift.

Material. Listeners' perception bias was determined with the help of missing fundamental stimuli, an idea that originated with Smoorenburg [3] who introduced a forced-choice task involving sequences of two complex tones. In the task, participants are presented with the 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 $F_n, F_{n+1}, ...F_{n+m-1}$, ($n$ 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 lowest partials (F_n). When the subsequent fundamentals F_0 are rising, the lowest partials F_n 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 ΔP_f by computing $\Delta P_f = (F^2_{k+1} - F^2_k) - (F^1_{k+1} - F^1_k)$ ($k \in \{n, n+1, ...n+m-2\}$) in order to estimate $F^2_0 - F^1_0$. Spectral listeners will rely on $\Delta P_{sp} = F^2_n - F^1_n$ to determine if the pitch is rising or falling. Figure 1 illustrates an ambiguous tone sequence for $n = 2$ and $m = 3$. The sequence depicted has a falling F_0 ($\Delta P_f < 0$) and a rising F_n ($\Delta P_{sp} > 0$).

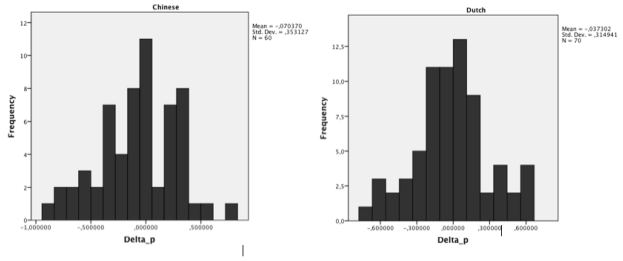

Figure 2: Distribution of δ_p, the measure of dominant listener mode, in the Chinese and Dutch group.

A total of 36 pairs of complex harmonic tones consisting of two, three or four harmonics were constructed following the procedure described in [4], including the addition of pink noise to minimize the effects of combination tones. Eighteen tone pairs were ambiguous, meaning that the second tone sequence would be judged as higher or lower than the first one depending on the participants listening mode. The remaining 18 tone pairs were unambiguous and were used as control stimuli. The tone pairs were offered in an order that was randomized per participant.

Procedure. All stimuli were presented using E-Prime (Psychology Software Tools, Inc., `www.pstnet.com`) and were heard over high quality headphones (Sennheiser Headset PC 320).

2.2. Results

We first compared the Dutch and the Chinese listeners with respect to the descriptive values for the proportion of correctly perceived non-ambiguous stimuli. The proportion correct served both as a measure of hearing quality and attention to the task. There was a difference between the two groups, with Dutch listeners performing slightly better (M = .898, SD = .096) than the Chinese listeners (M = .831, SD = .147), t(98.88) = 3.00, p = .003, equal variances not assumed (95% CI [0.022; 0.111]).

With respect to δ_p, the measure of dominant listener mode, the difference between the two groups was not significant, t(128) = 0.564, p = .574. The distribution for both groups was normal (Shapiro Wilk's tests of normality $p > .05$) and centered around 0, i.e. neither fundamental, nor spectral, see figure 2.2. The descriptive values for the Chinese listeners were $M = -0.070$, $SD = 0.353$ and thus comparable to the Dutch listeners, $M = -0.037$, $SD = 0.315$ (95% CI [-0.083; 0.149]).

3. Modelling study

The speech experienced throughout the development shapes perception. Language dependent perception implies that perception of pitch in listener tasks may be affected by native language. Our experimental results indicate that individuals raised in a non-tonal and tonal speech environment do not differ in their performances on listener tasks. This outcome suggests that the differential speech experience of non-tonal and tonal languages does not affect the perception of pitch in missing-fundamental tasks. In this section we verify this suggestion by means of a computational model of pitch perception proposed by Schwartz and Purves [18], henceforth referred to as the SP

model.

3.1. The Schwartz-Purves model of Pitch Perception

The basic premise of the SP model is that the main task of the auditory system is to infer what constitutes the most likely auditory source of an auditory stimulus. This probabilistic inference task can be approached from a Bayesian perspective (see [18]). The SP model is inspired by the straightforward idea that whenever confronted with an auditory stimulus, the brain can compare it to a vast number of stored memories of similar sound patterns and their associated meanings. For instance, the perceived pitch of a complex sound (e.g., a complex tone with a missing fundamental) corresponds simply to the pitch of the best-matching auditory representations stored in memory. In the SP model, determining the pitch of an input stimuls proceeds as follows. Given an input stimulus (of a small predefined length), the similarity with all voiced segments in a corpus is determined using cross-correlation. For each segment two values are registered: (1) the maximum matching value and (2) the pitch of the best-matching segment. The result can be represented as a scatter plot in which each point represents a pitch value (abscissa) and match value (ordinate). Pitch values of segments that are often well-matched to the stimulus yield a high peak. The height of the peak is proportional to the likelihood that the stimulus is assigned that pitch value. The SP model accounts for the perception of the missing fundamental by means of the large match that a missing-fundamental stimulus has with many stored sound segments that include the fundamental. Interpreting the SP model as a crude model of pitch perception in the brain, we can define a tonal and non-tonal version of the model and determine how both models respond to the missing-fundamental task.

We used a Chinese corpus (Mandarin Affective Speech Corpus, LDC2007S09) and an English corpus (TIMIT, LDC93S1), for creating such models. The corpora are considered to be reasonable proxies of the speech experience of an average tonal or non-tonal listener. As a consequence our tonal and non-tonal SP models represent average listeners of both language types. We selected all voiced segments by only accepting those segments that where classified as *voiced* with a probability of at least 95% by the PEFAC pitch tracker [19]. The selected segments were labelled with the pitch values estimated with the same tracker. To simulate our behavioral experiment, we presented both models with a missing-fundamental stimulus consisting of three partials, i.e., 450 Hz, 600 Hz, and 750 Hz for a duration of 50 ms (missing fundamental frequency $= 150$ Hz). We did not (have to) use a sequence of tones as we did in our behavioral experiment, because the SP model provides direct access to the "perceived" pitch.

Figure 3.1 presents the results obtained for the tonal and non-tonal versions, respectively. The plots show the maximum cross-correlation values (matching values) for the missing-fundamental stimulus and the pitch value of the best matching stored segment. Each dot is the result of cross-correlating the missing-fundamental stimulus to a voiced speech segment in our corpus. As can be seen, both the tonal and non-tonal versions of the SP model have the largest cross-correlations for a pitch of 150 Hz which corresponds to the pitch perceived by fundamental listeners. The presence of the largest peaks at this frequency suggest that the fundamental mode is the dominant listening mode for both Chinese and English. Interestingly, the second largest peak of the tonal model is located at the "spectral pitch" of 450 Hz, which is much less pronounced in the non-tonal model. This may be interpreted to imply that in tonal languages the spectral listening mode is more frequent than in non-tonal languages. The extent to which this is true should be determined in additional simulation studies.

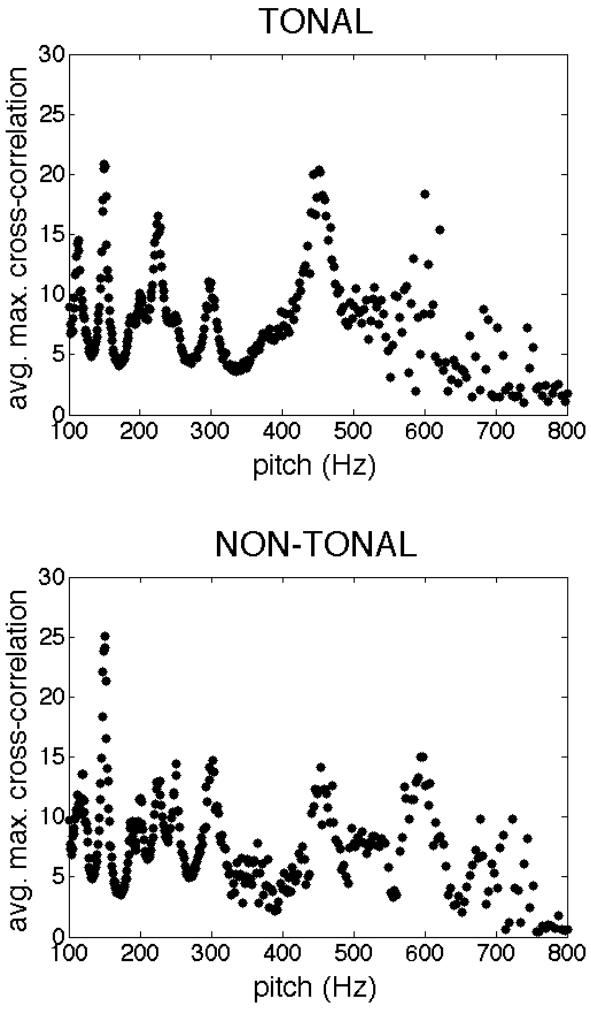

Figure 3: Top: Response of the tonal SP model to the missing-fundamental stimulus (450-600-750Hz). Each dot represents a comparison of the stimulus to a corpus segment. The coordinates correspond to the pitch value (abscissa) and the average cross-correlation value of the stimulus for all segments with that pitch value (ordinate). Bottom: Response of the non-tonal SP model.

4. Discussion and Conclusions

The two approaches used in the current study resulted in comparable outcomes. In the perceptual task, we controlled for the potential disturbing effect of combination tones. We found no differences in listener type distribution in groups of Dutch and Chinese listeners. Similarly, our computational investigation did not reveal a difference between the average English and Chinese model listeners. Both behave as fundamental listeners in that the fundamental frequency gives the largest response.

At a more detailed level, our computational study did reveal differences in pitch prominence for both models. Apparently, the non-tonal and tonal linguistic diets may give rise to

differences in the relative likelihoods of pitch values. As long as these pitch values do not give rise to the largest peaks, they do not represent the perceived pitch according to the decision rule of the SP model. However, a direct comparison of the results of our perceptual task and computational model is hampered by the fact that the former yielded a distribution of listener types, whereas the latter generated average listener types for both language types. A direct comparison would be facilitated by creating for each language multiple SP models, each of which would be created from random (but overlapping) subsets of the speech corpora. In this way, individual variations in the linguistic diet would be modelled. It is quite conceivable that these modelled individual variations would lead to SP models with slightly different pitch peak distributions resulting in spectral pitch perception. This is especially likely for the tonal SP model, because the second-largest peak is relatively large and corresponds to the perceived pitch of the spectral listener modus.

5. Acknowledgements

Part of this work was funded by a Chinese Research Council scholarship awarded to Yu Gu. We would like to thank the participants in our study and to Caixia Liu for her help with collecting the data for the perceptual task.

6. References

[1] H. von Helmholtz, *On the Sensations of Tone*. London: Longmans, 1885.

[2] A. Houtsma, "Musical pitch of two-tone complexes and predictions by modern pitch theories," *Journal of the Acoustical Society of America*, vol. 66, pp. 87–99, 1979.

[3] G. Smoorenburg, "Pitch perception of two-frequency stimuli," *Journal of the Acoustical Society of America*, vol. 48, pp. 924–942, 1970.

[4] V. Laguitton, L. Demany, C. Semal, and C. Liégeois-Chauvel, "Pitch perception: A difference between right- and left-handed listeners," *Neuropsychologia*, vol. 36, pp. 201–207, 1998.

[5] L. Rousseau, I. Peretz, C. Liégeois-Chauvel, L. Demany, C. Semal, and S. Larue, "Spectral and virtual pitch perception of complex tones: An opposite hemispheric lateralization?" *Brain and Cognition*, vol. 30, pp. 303–308, 1996.

[6] A. Seither-Preisler, L. Johnson, K. Krumbholz, A. Nobbe, R. Patterson, S. Seither, and B. Lütkenhöner, "Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 33, pp. 743–751, 2007.

[7] D. Ladd, R. Turnbull, C. Browne, C. Caldwell-Harris, L. Ganushchak, K. Swoboda, V. Woodfield, and D. Dediu, "Patterns of individual differences in the perception of missing-fundamental tones," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 39, 2013.

[8] P. Schneider, V. Sluming, N. Roberts, M. Scherg, R. Goebel, H. Specht, H. Dosch, S. Bleeck, C. Stippich, and A. Rupp, "Structural and functional asymmetry of lateral heschl's gyrus reflects pitch perception preference," *Nature Neuroscience*, vol. 8, pp. 1241–1247, 2005.

[9] P. Schneider, V. Sluming, N. Roberts, S. Bleeck, and A. Rupp, "Structural, functional and perceptual differences in heschl's gyrus and musical instrument preference," *Annals of the New York Academy of Sciences*, vol. 1060, pp. 387–394, 2005.

[10] P. Wong, C. Warrier, V. Penhune, A. Roy, A. Sadehh, T. Parrish, and R. Zatorre, "Volume of left heschl's gyrus and linguistic pitch learning," *Cerebral Cortex*, vol. 18, pp. 828–836, 2008.

[11] T. Griffiths, "Functional imaging of pitch analysis," *Annals of the New York Academy of Sciences*, vol. 999, pp. 40–49, 2003.

[12] V. Delvaux and A. Soquet, "The influence of ambient speech on adult speech productions through unintentional imitation." *Phonetica*, vol. 64, pp. 145–173, 2007.

[13] D. Dediu, "The role of genetic biases in shaping the correlations between languages and genes." *Journal of Theoretical Biology*, vol. 254(2), pp. 400–407, 2008.

[14] P. Wong, B. Chandrasekaran, and J. Zheng, "The derived allele of *aspm* is associated with lexical tone perception," *PLoS ONE*, vol. 7, no. 4, pp. 1–8, 2012.

[15] P. Wong and T. Perrachione, "Learning pitch patterns in lexical identification by native english-speaking adults," *Applied Psycholinguistics*, vol. 28, pp. 565–585, 2007.

[16] R. Plomp, "Detectability threshold for combination tones," *The Journal of the Acoustical Society of America*, vol. 47, pp. 1111–1123, 1965.

[17] E. Terhard, "Pitch, consonance and harmony," *Journal of the Acoustical Society America*, vol. 55, pp. 1061–1069, 1974.

[18] D. A. Schwartz and D. Purves, "Pitch is determined by naturally occurring periodic sounds." *Hearing Research*, vol. 194, no. 1-2, pp. 31–46, 2004.

[19] S. Gonzalez and M. Brookes, "A pitch estimation filter robust to high levels of noise (pefac)." in *European Signal Processing Conference (EUSIPCO 2011)*, 2011, pp. 451–455.