Psychoacoustic abilities as predictors of vocal emotion recognition

Eitan Globerson[1, 2], *Michal Lavidor*[1,3], *Ofer Golan*[3], *Liat Kishon-Rabin*[4], *Noam Amir*[4]

[1]Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat-Gan, Israel
[2]Jerusalem Academy of Music and Dance, Israel
[3]Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel
[4]Department of Communication Disorders, The Sackler Faculty of Medicine, Tel-Aviv University, Israel

Abstract

The mechanisms underlying vocal emotion recognition (VER) have been the subject of extensive research in the last decades. Evidence supporting a linkage between the level of musical background and vocal emotion recognition abilities was indicated in several studies, while others pointed to a linkage between Theory of Mind/emotional intelligence and VER. In the current paper we highlight pitch discrimination abilities as successful predictors of VER. Our results may have a significant impact on the assessment and rehabilitation of individuals suffering from deficient VER

Index Terms: Prosody, vocal emotion recognition, psychoacoustic thresholds, gliding tone threshold

1. Introduction

The term *prosody* refers to the acoustic attributes (such as intonation) of speech which can alter the meaning of an utterance. Prosody is an essential tool of human vocal communication, supporting the emotional and pragmatic messages conveyed in speech. Intact abilities to comprehend emotional and pragmatic prosodic cues in language are extremely vital for socio-emotional functioning. A great variability in prosodic perceptual abilities has been reported in the general population [1,2].

Deficiencies in receptive or productive prosodic abilities unaccompanied by other pathological symptoms are usually not considered abnormal. On the other hand, it is possible that even minor impairments in prosodic abilities may interfere with daily social interaction. This renders the study of prosodic abilities and the mechanisms supporting them a universal issue.

The search for mechanisms underlying Vocal Emotion Recognition (VER) has focused on both auditory and emotional abilities as possible candidates. Some studies have linked prosodic receptive skills to emotional abilities such as Emotional Intelligence (EI) [3] or Theory of Mind (ToM) [4]. Others have found a strong linkage between musical background and VER abilities [5-9]. Those supporting the EI/ToM-VER linkage argue that an impaired ability to comprehend the mental state of the other will hamper the ability to deduce one's mental state from the prosodic information conveyed in speech. Studies supporting the musical background-VER linkage demonstrated ample evidence for the superiority of musicians in VER tasks. In addition to those findings, Patel et al [9] have shown that individuals diagnosed as amusic (tone deaf) demonstrate inferior VER in speech. A possible explanation to the musical background-VER linkage is related to the findings that musicians demonstrate superior psychoacoustic abilities compared to non-musicians. Prior studies have reported, for example, superior frequency discrimination abilities in musicians [10,11]. In other words, it is possible that the reason for the superiority of musicians in detecting emotions in voice is not a direct result of their possible musical abilities, but rather, their superior psychoacoustic abilities.

Therefore, the goal of the present study was to investigate specific auditory mechanisms underlying VER, which could also be responsible for the music-VER linkage.

In order to examine our hypothesis we constructed a battery of tests employing both low-level psychoacoustic tests and an extensive emotional speech test. The psychoacoustic tests employed both steady and dynamic tone frequency discrimination tasks. These tasks contained discrimination assignments, in which the participants did not have to name a pitch direction, as well as tasks in which the participants had to name the pitch relationship between two steady tones, or the direction of a dynamic tone. Our purpose in constructing this experimental design was to examine the roles of both pitch discrimination and pitch direction naming in vocal emotion recognition.

2. Methods and participants

2.1 Participants

A group of 60 participants (30 male and 30 female participants) aged 20-36 years (mean: 25.3, std: 4.25) took part in the experiment. All participants were native Hebrew speakers with no known neurological or psychiatric conditions. Participants underwent screening audiometry at 0.5-4 KHz and demonstrated bilateral pure-tone air-conduction thresholds within normal limits. Participants had no previous experience in psychoacoustic testing.

2.2 Experimental design

Four psychoacoustic tasks and a four-block VER task were presented to the participants binaurally with Sennheiser HD-201 linear headphones and a line 6 ux1 external sound card. Background noise was assessed prior to the experiment with a sound level meter, to ensure a quiet working environment (below 40 DB). The order of the assignments was randomly modified between subjects to avoid an order effect. Participants were allowed to take as many breaks as needed between the various tasks. On average, participants took 150 minutes to complete all tasks (including breaks).

Psychoacoustic Tasks

The thresholds of the four psychoacoustic tasks were obtained using a 2 down 1 up adaptive staircase procedure, converging at a performance level of 70.7% [12]. The initial step size was 40 Hz. The step size was divided by 2 after each reversal, until a final step size of 1 Hz was reached. Assessment terminated after 10 reversals with the final step size. Thresholds were calculated using the arithmetic mean of the last 8 reversals.

All stimuli were 300 ms in duration, (including rise and fall ramps) with a 500 ms inter-stimulus gap.

All tests were preceded by a training session. Participants had to achieve 5 subsequent successes in the relevant tasks in order to begin the experiment.

1. Steady pure tone (PT) frequency discrimination tasks

a. *High/Low Discrimination Task* (HLDT): Two constant frequency pure tones were presented in each trial. Participants had to indicate which tone was higher.

b. *Oddball Paradigm Task* (OPT): a three-interval 2-alternative choice paradigm was employed. A fixed reference tone was followed by 2 other tones, one of which was a repetition of the reference and the other was different. Participants had to indicate which tone (2nd or 3rd) was different than the reference tone.

2. ***Gliding tone discrimination task*s**

The stimuli in this part of the experiment were pure tones with a linear, unidirectional change of frequency over time. Using a dedicated computer software (written in Matlab, 2007b), we evaluated the participants' Just Noticeable Difference (JND) values for pitch change in gliding tones (GT). Two pure tone GT tests were employed, as follows (see Figure 1 for an illustration of all psychoacoustic tasks).

a. *Gliding Tone Discrimination task* (GTD): a two-alternative forced-choice task, in which the participant was asked which one of two auditory stimuli was changing in pitch. Stimuli were either ascending or descending sinusoidal glides, 300 ms in length, with a center frequency of 1000 Hz. The slope of the gliding tone was gradually reduced, until the participant's JND was reached.

b. *Ascending/Descending Gliding Tone Discrimination task* (ADGTD): 1 interval 2 AFC task, in which the participant was asked to decide whether a gliding tone was rising or falling in frequency. As in the previous task, the slope of the gliding tone was gradually reduced, until the participant's JND was reached

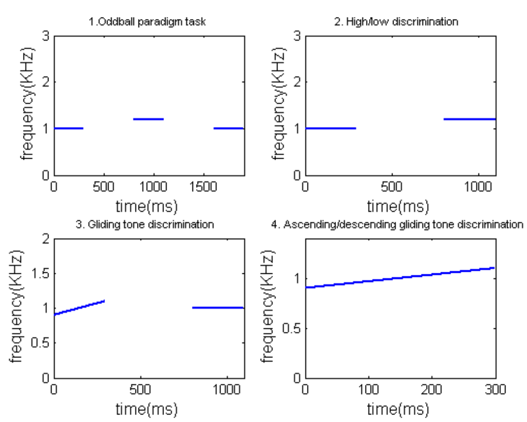

Figure 1: Illustration of the psychoacoustic tasks

***Vocal Emotion Recognition Task* (VERT)**:

All the stimuli for the VERT were recorded in a professional recording studio by four professional actors (two female, and two male actors). A total of 302 stimuli were included in the VERT. The stimuli were composed of monosyllabic utterances, nonsense words, neutral Hebrew words and neutral sentences, unrelated to the emotions expressed by the actors. Prior to the experimental process, the recorded stimuli were validated by a panel of 20 judges, thus ensuring that the emotional meaning intended by the actors was reliably expressed. All stimuli were normalized for identical RMS amplitude.

Participants heard the stimuli and had to decide which emotion was conveyed by the actors. The possible choices were: anger, fear, happiness, sadness, or neutral. The battery was divided into blocks, characterized by the nature of the utterances (monosyllabic utterances, words, etc.) The order of the blocks was randomized, as well as the order of the stimuli within the blocks. After naming the emotion, participants also had to rate the intensity of the emotion (except when "neutral" was selected), on a scale of three. The "neutral" option was not one of the designated emotions (i.e., the actors were not instructed to convey "neutral emotion"). This option was given to test the inclination of the participant to define non neutral utterances as neutral.

3. Results

A total of six scores were obtained for each participant, as follows.

a. The VERT score (in percent correct): the VERT score presented in our results is the arithmetic average of the scores in the individual sections of the VERT battery. The decision to use the average VERT score was established through factor analysis which confirmed that all the scores of the individual VERT blocks (characterized by their stimulus type: monosyllabic utterances, non words, words and sentences) have similar loadings on the general factor .

b. The "neutral score" (NS) :the percentage of stimuli rated as "neutral" by the participant in the voice emotion recognition task.

c. Four thresholds of the psychoacoustic tasks which were log-transformed (The employment of a logarithmic transformation of frequency discrimination thresholds is common practice in psychoacoustic literature [10]).

The group mean of each of these scores, as well as range and standard deviation are shown in Table 1.

Table 1: Range, mean and standard deviation of the results in each of the experimental tasks

task	Min.	Max.	Range	Mean	Std. Dev.
Vocal Emotion Recognition (percent)	60.60	92.05	31.46	79.07	7.41
Vocal Emotion Recognition Neutral Score (percent)	1.66	32.78	31.13	11.39	6.37
Oddball paradigm task (Hz)	1.00	63.50	62.50	9.43	10.29
High/low discrimination task (Hz)	.750	200.00	199.25	27.04	43.84
Ascending/descending gliding tone discrimination task (Hz)	2	199	197	23.87	36.14
Gliding tone discrimination task (Hz)	4.50	122.25	117.75	20.28	20.15

3.2 Bivariate Correlations

a. Correlations between psychoacoustic scores and VER

All psychoacoustic scores correlated significantly with VER scores: HLDT score ($r=-.506$, $p<.001$), OPT score ($r=-.255$, $p<.05$), ADGTD score ($r=-.490$, $p<.001$, GTD ($r=-.391$, $p<.005$). All psychoacoustic scores had a negative relationship with emotion recognition abilities, such that better psychoacoustic abilities corresponded to better emotion recognition abilities (the negative correlation is due to the fact that lower thresholds correspond to better abilities). Note that correlation coefficients were higher for the psychoacoustic tasks that included pitch direction naming.

b. Correlations between psychoacoustic scores and neutral scores (NS)

Three psychoacoustic scores correlated significantly with NS: HLDT score ($r=.376$, $p<.005$), ADGTD score ($r=.543$, $p<.001$), GTD ($r=.361$, $p<.005$). Psychoacoustic scores had a positive correlation with NS, such that better psychoacoustic abilities corresponded to a smaller tendency to define a stimulus as neutral (figure 2). Note that also in the case of the NS, correlation coefficients were higher for the psychoacoustic tasks that included pitch direction naming

3.3 Stepwise Regressions

a. Psychoacoustic scores as predictors of VER

A stepwise regression was performed with all four psychoacoustic scores as predictors of the VERT score. The only two psychoacoustic scores to enter the analysis (in this order) were the HLDT score ($\beta=-.506$, $p<.001$) and ADGDT ($\beta=-.292$, $p<.05$) Together these scores explain 31.0% of the variance in the VER scores [$R=.557$, $F(2,57)=12.806$, $p<0.001$]. Again, it is important to mention that the significant psychoacoustic predictors of VER abilities are those tasks in which the participants had to name a direction of a steady or dynamic pitch interval. Once these predictors entered the analysis, contribution of the remainder of the pitch discrimination tasks became non-significant.

b. Psychoacoustic scores as predictors of NS

A stepwise regression was performed with all four psychoacoustic scores as predictors of NS. The only psyochoacoustic score to enter the analysis was the ADGDT score, which explained 29.5% of the variance in NS. As in the previous case, the only significant psychoacoustic predictor of NS was a task in which the participants had to name the direction of the pitch contour.

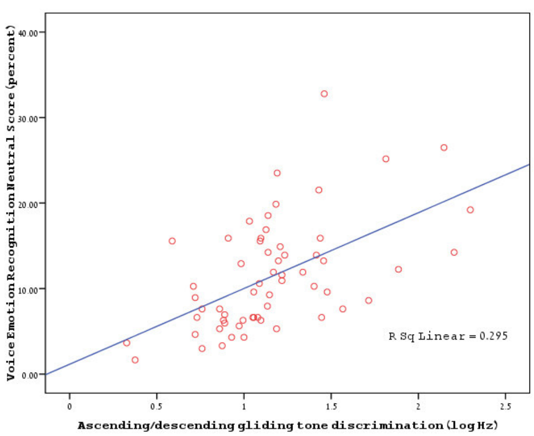

Figure 2: Vocal Emotion Recognition Task Neutral Score (NS) plotted against *Ascending/Descending Gliding Tone Discrimination* (ADGDT) thresholds

4. Discussion

This study was designed to investigate the possible association between pitch discrimination abilities and Vocal Emotion Recognition (VER). Our results demonstrate a significant correlation between these two parameters. We found that improved psychoacoustic abilities correlate with improved VER, while poor psychoacoustic abilities correlate with the participants' tendency to define emotional speech as neutral.

To our knowledge, these results are first time evidence demonstrating the importance of psychoacoustic abilities in VER in the general population. Prior studies have demonstrated a strong linkage between psychoacoustic abilities and speech recognition [13,14]. Our results point to an additional direction, namely, that psychophysical abilities can influence emotional vocal communication, and not only speech recognition.

4.1 The extra benefit of psychoacoustics

An obvious advantage in using psychoacoustic tests as predictors of VER is their relative short duration. The two tests which proved as successful predictors of VER abilities in our experimental battery require only about 5 minutes. This may prove useful when using psychoacoustics as a future diagnostic tool.

4.2 Discrimination vs. direction naming tasks

The psychoacoustic tasks employed in our study could be viewed as belonging to two distinct categories: the pitch *discrimination* category, and the pitch *direction naming* category. Our analysis of the results reveal that the pitch direction naming tasks were found to be better predictors of VER than the pitch discrimination tasks. One should keep in mind that the pitch direction naming tasks are also discrimination tasks, since the stimuli are adaptively altered until reaching the threshold beyond which the participant cannot perform the pitch direction naming task. It should also be noted that the scores in the discrimination tasks also correlated significantly with the VERT scores, although not as strongly as the scores in the direction naming category. Hence, both discrimination and pitch direction naming abilities may be vital for the recognition of emotion in prosody.

The existence of distinct brain mechanisms underlying pitch discrimination and pitch naming tasks has been described in previous studies [15]. Our results imply that both mechanisms are essential for processing prosodic information conveyed in speech.

4.3 Future perspectives

Our results, as far as we know, are the first to indicate that emotion recognition abilities correlate to basic psychoacoustic aptitudes. There could be significant future applications to these results. Prior studies have shown a great variability in psychoacoustic abilities in the general population. Other studies, however, have demonstrated a great effect of training on the possibility of improving these abilities [16]. Taken together, it would appear that such interventions in early childhood may substantially ameliorate the emotion recognition abilities of children in the early stages of development, thereby promoting their emotional communicational skills.

In addition, our results may pave the way to a more refined search for basic auditory abilities underlying vocal emotion perception. These results demonstrate the importance of both pitch discrimination, and pitch direction naming abilities in VER. Our findings call for further investigation, possibly employing brain imaging techniques.

Finally, our paradigm may be useful in the investigation of the mechanisms underlying the poor VER abilities found in various neuro-psychiatric disorders, such as autism [17] or schizophrenia [18].

5. Conclusions

Our study presents preliminary evidence for the role of psychoacoustic abilities in VER. These results may pave the way to a better understanding of the brain mechanisms supporting VER.

6. Acknowledgements

This study was supported by the NARSAD independent investigator award and by grant no. 474/06 from the Israel Science Foundation awarded to M. Lavidor.
We would also like to thank all participants in the tests for their time and effort.

7. References

[1] Sobin, C., Alpert, M. (1999). Emotion in Speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research*, 28(4), 347-365

[2] Hammerschmidt, K., Juergens, U. (2007). Acoustical correlates of affective prosody. *Journal of Voice*, 21(5), 531-540.

[3] Trimmer, C.G., Cuddy, L.L. (2008) .Emotional Intelligence, Not Music Training, Predicts Recognition of Emotional Speech Prosody. *Emotion*, 8 (6), 838–849

[4] Baron-Cohen, S., Tager-Flusberg, H., Cohen, D. J. (2000). Understanding other minds: perspectives from developmental cognitive neuroscience (2nd ed.). *Oxford: Oxford University Press*.

[5] Ayotte, J., Peretz, I., Hyde, K. (2002). Congenital amusia. *Brain, 125,*238–251.

[6] Nicholson, K. G., Baum, S., Cuddy, L. L., Munhall, K. G. (2002). A case of impaired auditory and visual speech prosody perception after right hemisphere damage. *Neurocase, 8,* 314–322

[7] Nicholson, K. G., Baum, S., Kilgour, A., Koh, C. K., Munhall, K. G., & Cuddy, L. L. (2003). Impaired processing of prosodic and musical patterns after right hemisphere damage. *Brain and Cognition, 52,* 382–389.

[8] Patel, A. D., Foxton, J. M., Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition, 59,* 310–313.

[9] Patel, A.D., Wong, M., Foxton, J., Lochy, A., Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (Congenital Amusia). *Music Perception*, 25(4), 357–368.

[10] Micheyl C, Delhommeau K, Perrot X, Oxenham AJ (2006) Influence of musical and psychoacoustical training on pitch discrimination. Hear Res 219(1–2): 36–47.

[11] Kishon-Rabin, L., Amir, O., Vexler, Y., Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than nonmusicians? *Journal of Basic and Clinical Physiology and Pharmacology*, 12,125–143.

[12] Levitt , H. (1971). Transformed up–down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, 467-477.

[13] Tallal, P. (2000). The science of literacy: From the laboratory to the classroom. *PNAS*, 97, 6.

[14] Kishon-Rabin, L, Segal, O., Algom, D.(2009). Associations and Dissociations between Psychoacoustic Abilities and Speech Perception in Children with Severe-to-Profound Hearing Loss. *Journal of Speech, Language, and Hearing Research*, 52, 956-972.

[15] Johnsrude, I., Penhume, V., Zatorre, R. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain,* 123**,** 155–163.

[16] Wright, B.A. (2001). Why and how we study human learning on basic auditory tasks. *Audiol Neurootol* ,6:207–210

[17] Golan, O., Baron-Cohen, S., Hill, J. J., Rutherford, M. D. (2007). The 'Reading the Mind in the Voice' test - Revised: A study of complex emotion recognition in adults with and without Autism Spectrum Conditions. *Journal of Autism and Developmental Disorders*, 37(6), 1096-1106.

[18] Murphy, D., Cutting, J. (1990). Prosodic comprehension and expression in schizophrenia. *Journal of Neurology, Neurosurgery, and Psychiatry*, 53, 727-730.