ASHA American Speech-Language-Hearing Association

JSLHR

Technical Report

A Cross-Linguistic Validation of the Test for Rating Emotions in Speech: Acoustic Analyses of Emotional Sentences in English, German, and Hebrew

Micalle Carl,[a] Michal Icht,[a] and Boaz M. Ben-David[b,c,d]

[a] Department of Communication Disorders, Ariel University, Israel [b] Baruch Ivcher School of Psychology, Reichman University (IDC) Herzliya, Israel [c] Department of Speech-Language Pathology, University of Toronto, Ontario, Canada [d] Toronto Rehabilitation Institute, University Health Network (UHN), Ontario, Canada

ARTICLE INFO

Article History:
Received April 9, 2021
Revision received July 17, 2021
Accepted November 23, 2021

Editor-in-Chief: Bharath Chandrasekaran
Editor: Chao-Yang Lee

https://doi.org/10.1044/2021_JSLHR-21-00205

ABSTRACT

Purpose: The Test for Rating Emotions in Speech (T-RES) has been developed in order to assess the processing of emotions in spoken language. In this tool, spoken sentences, which are composed of emotional content (anger, happiness, sadness, and neutral) in both semantics and prosody in different combinations, are rated by listeners. To date, English, German, and Hebrew versions have been developed, as well as online versions, iT-RES, to adapt to COVID-19 social restrictions. Since the perception of spoken emotions may be affected by linguistic (and cultural) variables, it is important to compare the acoustic characteristics of the stimuli within and between languages. The goal of the current report was to provide cross-linguistic acoustic validation of the T-RES.
Method: T-RES sentences in the aforementioned languages were acoustically analyzed in terms of mean F0, F0 range, and speech rate to obtain profiles of acoustic parameters for different emotions.
Results: Significant within-language discriminability of prosodic emotions was found, for both mean F0 and speech rate. Similarly, these measures were associated with comparable patterns of prosodic emotions for each of the tested languages and emotional ratings.
Conclusions: The results demonstrate the lack of dependence of prosody and semantics within the T-RES stimuli. These findings illustrate the listeners' ability to clearly distinguish between the different prosodic emotions in each language, providing a cross-linguistic validation of the T-RES and iT-RES.

The processing of emotions in spoken language plays an important role in daily interpersonal interactions (Ben-David et al., 2013; Ben-David, Ben-Itzchak, et al., 2020). Since many real-life emotional situations occur in a social-interactive context, in which the emotions are being initiated by other persons, listeners must infer emotions from various affective cues (Banse & Scherer, 1996). When a listener does not fully comprehend the emotion conveyed by the speaker, miscommunication arises, with possible negative implications for the quality of social interactions (Hudepohl et al., 2015; Icht, Wiznitser Ressis-tal, & Lotan, 2021).

The perception of spoken emotional cues involves the processing of data from several sensory modalities, including (but not limited to) visual and auditory information. In the absence of visual cues, such as during a phone conversation, the ability to derive emotional meaning is dependent on how it is delivered in two auditory speech channels—the semantic channel (the meaning of the words) and the prosodic channel (tone and intonation of voice; Leshem et al., 2020).

To understand the complex ability to process spoken emotions, the Test for Rating Emotions in Speech (T-RES; Ben-David et al., 2016) has been introduced. In this test, participants listen to sentences that present emotional,

Correspondence to Boaz M. Ben-David: boaz.ben.david@idc.ac.il.
Disclosure: *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

semantic, and prosodic content in different combinations, congruent or incongruent. In three separate tasks, listeners are asked to rate the extent to which they agree that a sentence conveys a predefined emotion (anger, happiness, sadness, and neutrality), on a Likert scale of 1 (*strongly disagree*) to 6 (*strongly agree*), while focusing on either the semantic or the prosodic channel, or on both. No feedback is provided throughout the experiment, as there are no "right" and "wrong" answers, rather, the T-RES gauges the listener's subjective perception of emotions. As a result, the performance on the T-RES's three tasks directly test three distinct components of emotional speech processing: (a) identification of emotions in the tone of speech (prosody) and semantics, (b) selective attention: focusing on one while ignoring the other channel, and (c) integration of the content of prosody and semantics: processing the sentence as a whole.

The T-RES has been extensively used to assess processing of emotional speech in different populations and languages. These include healthy young adults (Ben-David et al., 2016), older adults (Ben-David et al., 2019), people with tinnitus (Oron et al., 2020), undergraduates with high functioning autism spectrum disorder (ASD; Ben-David, Ben-Itzchak, et al., 2020), individuals with forensic schizophrenia (Leshem et al., 2020), and cochlear implant users (Taitelbaum-Swead et al., 2022). The original English version of this tool (Ben-David et al., 2016) has been translated and adapted to German (Defren et al., 2018) and to Hebrew (Shakuf et al., 2016).

Recently, in response to COVID-19 challenges, an online version (iT-RES; a remote adaptation) of the T-RES was developed (Ben-David, Mentzel, et al., 2020) in the three languages, freely available for use (http://www.canlab.idc.ac.il/itres). The Hebrew remote version (tested at the participants' home) was validated against a traditional lab version, conducted in a sound-attenuated booth. However, a cross-linguistic validation has not yet been conducted. Since the perception of spoken emotions may be affected by linguistic (and cultural) factors (Ekman et al., 1987), such analysis is needed in order to compare T-RES findings in different languages, and to generalize their results. Filling this gap in the literature, the current report compares the acoustic characteristics of the T-RES emotional spoken sentences, within and between languages, focusing on the methodological aspects of the stimuli analysis and conclusions drawn from these results.

Within-Language Differences: Specific Vocal Expression Patterns for Different Emotions

The literature demonstrates that different emotions can be expressed and recognized based on certain acoustic variables that include: (a) fundamental frequency (F0; in Hz): the frequency of the vocal folds' vibration, perceived as vocal pitch, and more specifically, F0 level, range, and contour; (b) intensity: the amount of vocal energy (amplitude; in dB), perceived as loudness; and (c) temporal characteristics, such as speech rate and pausing (Borden & Harris, 1984; Scherer, 1989). For example, anger and happiness are typically characterized by high mean pitch and high mean voice intensity, whereas sadness is characterized by low mean pitch and low mean voice intensity (Juslin & Laukka, 2003). Happiness and anger expressions typically have large pitch variability and a fast speech rate, whereas sadness expressions have small pitch variability and a slow speech rate (Banse & Scherer, 1996; Leitman et al., 2010). As some vocal emotions share similar acoustic characteristics, failure to use specific vocal cues may reduce the accuracy of identification and may lead to confusion between the different emotions (Leitman et al., 2010).

The T-RES stimuli, recorded spoken emotional sentences (Ben-David et al., 2011), present emotional prosodies of anger, happiness, sadness, and neutrality (Ben-David et al., 2013) produced by professional actresses. To validate the findings of the T-RES studies, it is important (a) to confirm that these prosodic emotions are indeed acoustically different, (b) to identify which of the acoustic measures distinguishes the different prosodic emotions within each language, and (c) to ensure that the semantic content of the sentences does not affect the prosodic acoustic characteristics.

Between-Language Differences: Vocal Expression Patterns in Different Languages

The literature suggests the universality (vs. cultural relativity) of emotional expressions, as they possess invariant characteristics (Pell et al., 2009). Many of the acoustic parameters involved in emotion-specific vocal profiles have been associated with emotion-specific physiological (phonatory and articulatory) characteristics (Scherer, 1986), suggesting that prosodic identification may be independent of learning or culture. Indeed, there are several studies indicating that members of one culture correctly identify the meaning of the prosodic emotional expressions in another culture (e.g., Pell et al., 2009; Pell & Skorup, 2008; Scherer et al., 2001). Other studies, however, point to differences in emotion expressions across cultures, due to socio-cultural dimensions, such as cultural norms (Ekman et al., 1987; Elfenbein et al., 2007). In fact, many of the studies postulating cultural similarities, still report an advantage for identifying vocal emotion expressions in the listener's native language (see also Elfenbein & Ambady, 2002).

The T-RES paradigm presents a unique purview on the universality/culture specificity of prosodic emotions,

given the languages used (English, German, and Hebrew). Both English and German are Indo-European (Germanic) languages, whereas Hebrew is Afro-Asiatic (Semitic). These differences are reflected in the phonetic inventories and intonation patterns (Abu-Rabia, 2001; Hurley, 1992). These three T-RES versions were recorded and validated in different counties (Canada, Germany, and Israel) that differ on various cultural attributes (e.g., collectivism vs. individualism; Hofstede, 2001). Notwithstanding these cross-linguistic and cross-cultural differences, we expected to find similar overall patterns of the effect of emotional prosodies on acoustic features, supporting the universality of the tool.

The Current Report

The current report aims to validate the use of the T-RES in three languages with an acoustic analysis of the tests' stimuli. Two main sets of hypotheses stem from the aforementioned literature, as related to within-language and between-languages differences. Within-language: We expected to find large between-emotion differences in acoustic characteristics of emotional prosodies in each linguistic set and to find that prosodic characteristics were not affected by the semantic content of the sentence. Between-languages: It was hypothesized that results of acoustic measures (namely, mean F0, F0 range, and speech rate) will demonstrate similar acoustic trends for the tested emotions across languages.

Materials and Method

Speech Stimuli

The English, German, and Hebrew versions of the T-RES (Ben-David et al., 2013; Defren et al., 2018; Shakuf et al., 2016) were used, with the following emotions: anger, happiness, sadness, and neutrality. T-RES stimuli consist of 32 spoken sentences in which each semantic emotion is represented twice in each of the tested prosodies (with eight spoken sentences per prosodic emotion). Sentences were composed of both congruent and incongruent semantic and prosodic emotions. For example, the semantically happy sentence "I won an award" is presented with a congruent happy prosody while the sentence "Congratulations, you're hired" is presented with an incongruent angry prosody. These sentences were recorded by native speakers of English, German, and Hebrew, who are all professional actresses, using the four different prosodies. Recordings were conducted in recording studios, digitized (16-bit) at a sampling rate of 44 kHz. Digital audio files were equated with respect to their root-mean-square amplitude and duration. When constructing the separate T-RES linguistic versions, 750 recorded versions were created in each language. Out of which, the subset of spoken sentences was chosen based on perceived high quality of prosodic information. In this process, outliers with respect to extreme acoustic characteristics, or poor representations of intended emotions, were removed (for details, see Ben-David et al., 2013). For a full description of the characteristics of the spoken sentences, see the research of Ben-David et al. (2019) and Ben-David, Ben-Itzchak, et al., (2020). The original audio files are available at http://www.canlab.idc.ac.il/itres.

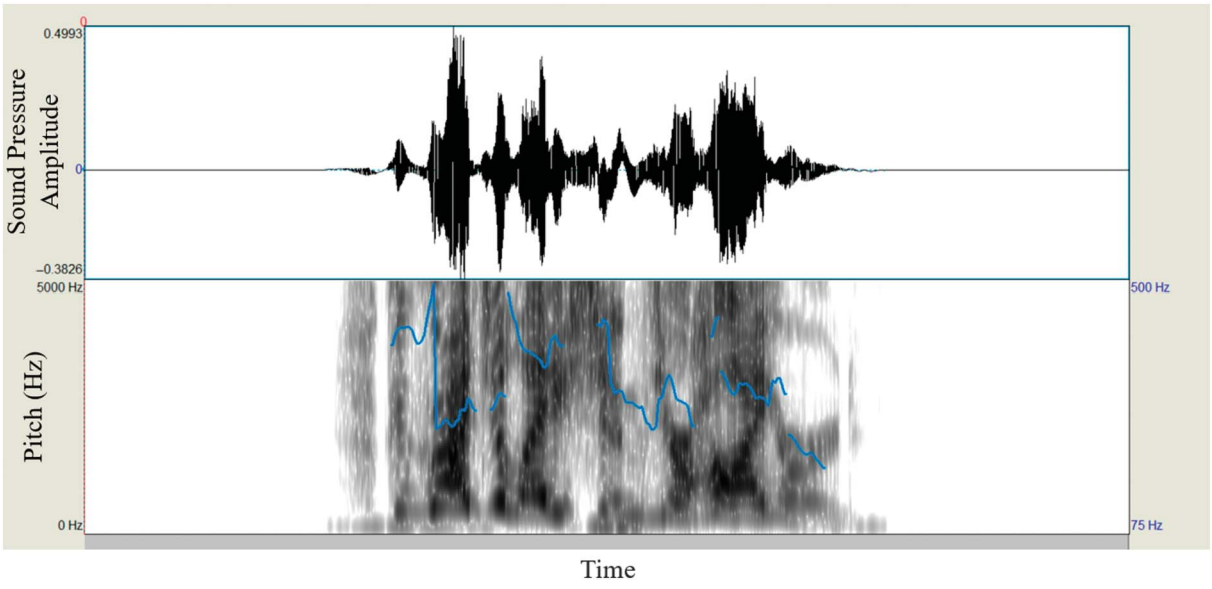
Acoustic Analysis

Acoustic analyses of the sentences were performed using Praat software, Version 6.1.07 (Boersma & Weenink, 2019). An example of the voice recordings used is shown in Figure 1. The analyses were conducted by the first and second authors, who are experienced speech-language pathologists trained in acoustic analysis.

Three primary measures were calculated for comparisons between languages (as reported by Bänziger & Scherer, 2005; Pell et al., 2009): (a) average F0 value of the sentence (normalized Hz)—representing the average fundamental frequency of the entire utterance; (b) F0 range (normalized Hz)—representing the range between the maximum and minimum F0 values, calculated from the entire utterance; and (c) speech rate—calculated as the number of syllables per phonatory duration of the utterances (number of syllables/utterance duration), as counted by native speakers of the respective languages. The duration of the utterance was calculated as the beginning to the end of speech production, whether phonation or voiceless consonant, measured in seconds. Mean, minimum, and maximum F0 values were obtained by means of a Praat script, in which minimum pitch was set to 100 Hz and maximum pitch to 600 Hz. The standard, autocorrelation algorithm in Praat was used to extract F0 from the sentence stimuli. It should be noted that intensity measures were not included, as the sentences were equalized with respect to intensity, as detailed above.

In order to compare between speakers of each language, pitch results were normalized (as conducted in Pell et al., 2009), as follows. The average minimum pitch for neutral sentences of each language was calculated and termed the *resting frequency* for that language. The normalized mean F0 value was calculated by subtracting the resting frequency from the mean F0 of each utterance, and then dividing each by the resting frequency, as detailed in Equation 1. Normalization was also conducted for minimum and maximum pitch values for each sentence using an identical calculation, with the exception of

Figure 1. Acoustic analysis of sample English Test for Rating Emotions in Speech sentence: "Congratulations, you're hired." Semantic emotion: Happiness, Prosody emotion: Anger. The waveform of the utterance is found in the upper portion of the figure, while the spectrographic display is depicted in the lower rectangle. The *x*-axis represents time, and the *y*-axis represents sound pressure amplitude for the waveform and pitch (in Hz) for the spectrogram. The solid line in the lower section represents the pitch contour across the sentence.



minimum or maximum F0 values in the place of mean F0 (Pell et al., 2009).

$$Normalized\ Mean\ F0 = \frac{Mean\ F0 - Resting\ Frequency}{Resting\ Frequency} \quad (1)$$

Statistical Analysis

All statistical analyses were conducted in R (R Core Team, 2020). To test for normality of the data, a measure of skewness was calculated for each acoustic variable (mean F0, F0 range, and speech rate), using the "Moments" package in R (Komsta & Novomestky, 2015). Skewness values of 0.57, −0.33, and 0.48 across variables, respectively, were reported, indicating no severe violation of skewness of the data that may affect analyses, as based on a threshold of ±1.75 (Blanca et al., 2013).

Mixed-model analyses of variance (ANOVAs) were conducted, for each of the acoustic measures (mean F0, F0 range, and speech rate), separately, using the *aov* function in *car* library of R (Fox & Weisberg, 2019). Independent variables included prosodic emotion (X4: anger, happiness, sadness, neutrality) and congruency (X2: semantic content congruent or incongruent with the prosody) as within-subject variables, and language (X3: Hebrew, German, and English), as a between-subject variable. To further test for potential differences between languages and/or emotions, post hoc analyses were conducted for incongruent sentences alone, using a dedicated ANOVA, for each of the acoustic measures. Due to lack of congruent sentences for the neutral emotion, as they were removed from the original T-RES stimuli (Ben-David et al., 2016, 2019; Defren et al., 2018), post hoc analyses were conducted for incongruent trials alone. Estimated marginal means (Prosody alone and Prosody*Language) with Tukey-adjusted pairwise comparisons were calculated for cross- and within-language comparisons, using the *emmeans* package in R (Lenth, 2019). For an estimate of effect size, partial eta squared (η_p^2) was used, based on the following ranges for reported values: small effect size = 0.01; medium effect size = 0.06; large effect size = 0.14 (Richardson, 2011). The alpha level for significance of statistical analyses was specified at $\alpha = 0.05$.

A post hoc estimate of observed power was calculated for statistically significant findings, using the G*Power software (Version 3.1; Faul et al., 2009) as follows. An effect size (Cohen's *f*) was calculated from the smallest η_p^2 value among significant findings, given that all significant *p* values were $p < .001$. The calculated effect size ($f = 0.55$) was used along with input values of α error probability ($\alpha = 0.01$), total sample size ($n = 90$), numerator degrees of freedom (6), and number of groups (3) to determine the estimate of minimal power. This estimate was calculated as 0.925, indicating strong statistical power to reject a false null hypothesis.

Results

A summary of acoustic results is presented in Table 1, separately by language and emotional prosody. Table 1 also includes the average perceptual ratings of the various emotions, as judged by participants (typically developing young adults) in previous T-RES studies (Ben-David et al., 2016, 2019; Defren et al., 2018; Shakuf et al., 2022). The perceptual ratings illustrate that prosodic sentences were indeed good representations of their intended prosodic emotional category, with extremely high ratings for each of the emotions, anger, happiness, and sadness (> 5.45 out of 6), relative to the neutrality sentences (≤ 2 out of 6), across all three languages.

Table 1. Means and standard deviations for each acoustic measure, across both languages and emotional prosodies, as well as emotional ratings of these sentences by native speakers of the language.

Emotion (prosody)	Measure	Hebrew	German	English
Anger	Mean F0	2.47 (0.41)	1.58 (0.5)	1.56 (0.22)
	Range F0	3.21 (0.5)	2.62 (0.79)	2.65 (0.40)
	Speech rate	4.95 (1.18)	4.72 (0.75)	4.04 (0.47)
	Emotional rating	5.85 (.48)	5.73 (.41)	5.53 (.52)
Happiness	Mean F0	2.11 (0.66)	1.5 (0.21)	1.54 (0.36)
	Range F0	2.41 (0.71)	3.12 (0.73)	2.34 (0.39)
	Speech rate	4.14 (0.5)	6.16 (0.7)	3.53 (0.64)
	Emotional rating	5.47 (.57)	5.77 (.39)	5.60 (.53)
Sadness	Mean F0	1.33 (0.34)	1.09 (0.22)	1.27 (0.48)
	Range F0	2.85 (1.09)	3.32 (0.62)	2.68 (0.72)
	Speech rate	3.77 (0.43)	4.58 (0.44)	3.1 (0.54)
	Emotional rating	5.67 (.55)	5.61 (.46)	5.60 (.54)
Neutrality	Mean F0	0.53 (0.1)	0.61 (0.12)	0.41 (0.15)
	Range F0	1.53 (1.08)	1.73 (1.07)	2.38 (0.83)
	Speech rate	4.41 (0.33)	5.35 (0.66)	3.55 (0.62)
	Emotional rating	1.55 (.58)	1.66 (.64)	2.01 (.74)

Note. Mean and range of F0 are presented in normalized Hz, speech rate is presented in syllables per second, and emotional rating* is presented on a scale of 1 (*unemotional*) to 6 (*highly emotional*).

*Data were taken from the following Test for Rating Emotions in Speech (T-RES) studies, with native speakers of the respective languages; Hebrew: Ben-David et al. (2019; n = 40), English: Ben-David et al. (2016; n = 80), and German: Defren et al. (2018) and Shakuf et al. (2022; by personal communication, n = 80). For example, average ratings for anger were taken from the scale "How much do you agree that the speaker is *angry*? From 1 – strongly disagree to 6 – strongly agree" for anger-prosody sentences. For neutrality rating, as in the original T-RES paradigm no scale directly assesses neutrality, the averages of anger ratings, sadness ratings, and happiness ratings for neutral prosody sentences were used. Thus, the lower the number, the more neutral (unemotional) the emotional rating.

F0 Mean (Normalized)

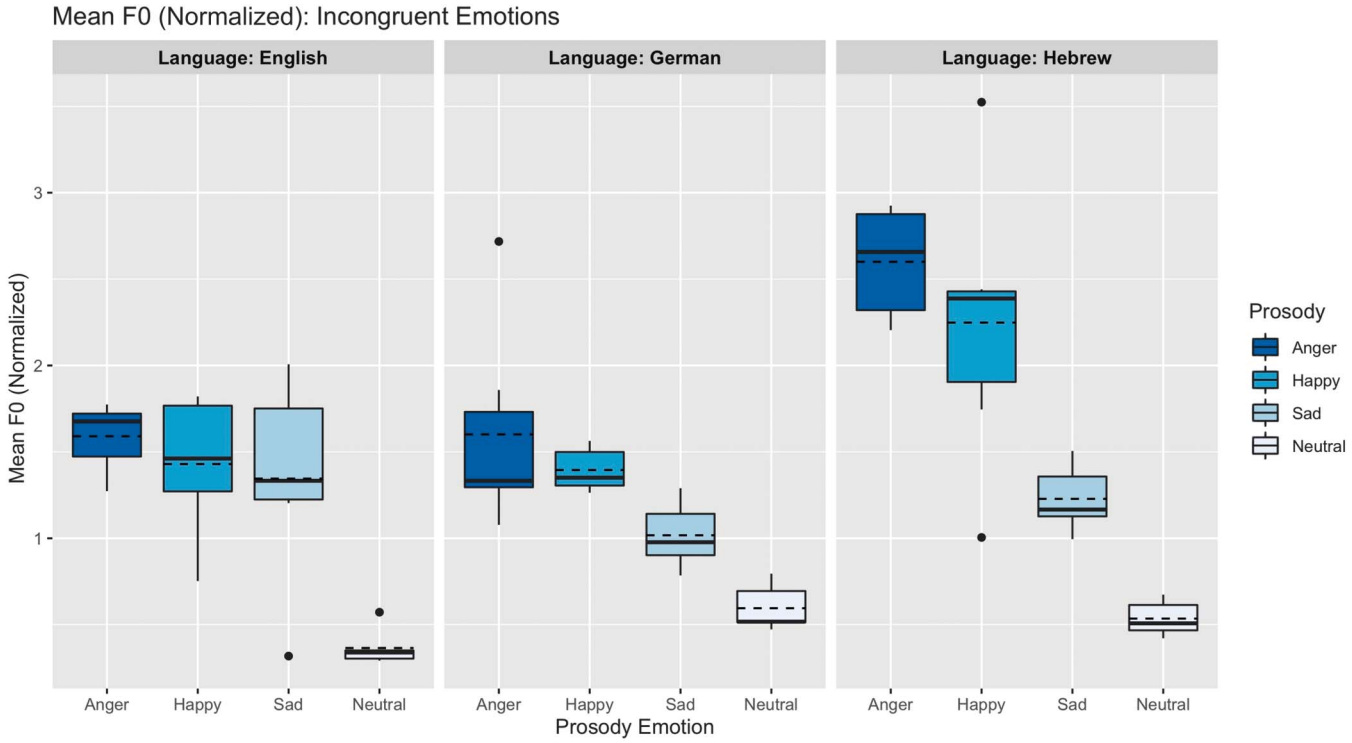
Results of the main analysis conducted for mean F0 demonstrated significant main effects for prosodic emotion, $F(3, 69) = 57.072$, $p < .001$, $\eta_p^2 = .702$, and language, $F(2, 69) = 16.099$, $p < .001$, $\eta_p^2 = .318$, as well as a significant interaction between the two, $F(6, 69) = 4.724$, $p < .001$, $\eta_p^2 = .297$. However, no significant main effect was found for congruency, $F(1, 69) = 0.024$, $p = .876$, nor did it lead to a significant interaction with language, $F(2, 69) = 0.429$, $p = .653$, with prosodic emotion, $F(2, 69) = 0.318$, $p = .729$, or a significant interaction of all three, $F(4, 69) = 0.524$, $p = .718$. In other words, mean F0 was affected by the prosodic emotion (in the direction of anger > happiness > sadness > neutrality), and the tested language (see Table 1). However, the semantic emotional content of the sentences, whether congruent or incongruent with the prosodic emotion, did not affect mean F0. For an illustration, the semantically happy sentence "I won the lottery" spoken with a congruent, happy, prosody, or with an incongruent, sad, prosody was produced with similar mean F0 values across languages.

Post hoc analyses of incongruent sentences indicated that across languages, mean F0 was affected by the prosodic emotion, with lowest values for neutrality followed by sadness, and with happiness and anger characterized by higher values. Contrasts of emotions across all three languages were significant for neutrality, as compared to all other three emotions (anger vs. neutrality, t.ratio = 10.881, $p < .001$; happiness vs. neutrality, t.ratio = 9.058, $p < .001$; sadness vs. neutrality, t.ratio = 5.305, $p < .001$); and for sadness, as compared with anger (t.ratio = 5.576, $p < .001$) and with happiness (t.ratio = 3.753; $p < .01$). Within-language contrasts revealed similarity in trends, with significant differences ($p \leq .01$) between neutrality and all other emotions in both English and Hebrew, and between neutrality and both anger and happiness in German ($p < .01$). Figure 2 provides a visual illustration of this analysis.

F0 Range (Normalized)

Results of the main analysis conducted for F0 range demonstrated a significant main effect for prosody, $F(3, 69) = 7.045$, $p < .001$, $\eta_p^2 = .233$. No further main effects were found to be significant: language, $F(2, 69) = 0.665$, $p = .518$, congruency, $F(1, 69) = 0.232$, $p = .631$, nor any of the interactions between them: prosody and language, $F(6, 69) = 1.596$, $p = .161$, prosody and congruency, $F(2, 69) = 0.653$, $p = .524$, language and congruency, $F(2, 69) = 1.351$, $p = .266$, nor the interaction between all three main effects, $F(4, 69) = 0.867$, $p = .488$. Taken together, it appears that range of F0 was affected by the prosodic emotion alone, whereas factors of

Figure 2. Box plots of mean F0 values (normalized, Hz), for incongruent sentences alone, across target prosodies and tested languages. The lower edge of the box represents the 25th quartile, the upper edge represents the 75th quartile (between them, the interquartile range; IQR), the solid line represents the median value, and the dotted line represents the mean value. The whiskers represent the highest and lowest values within 1.5 times the IQR. The solid dots represent outlying values (larger than 1.5 times the IQR).



language and congruency had no impact. In other words, F0 range varied by emotional prosody alone, without a significant difference between the three languages.

Post hoc analyses of incongruent sentences across languages demonstrated that neutrality was characterized by the smallest F0 range, as compared to all other three emotions (neutrality vs. anger, t.ratio = 3.467, $p < .01$; neutrality vs. happiness, t.ratio = 2.819, $p < .05$; neutrality vs. sadness, t.ratio = 4.066, $p < .05$). Within-language comparisons revealed significant contrasts between neutrality and sadness in both German (t.ratio = 3.236, $p < .05$) and Hebrew (t.ratio = 3.409, $p < .01$), and between neutrality and anger in Hebrew (t.ratio = 3.222, $p < .05$). No further effects were noted.

Speech Rate

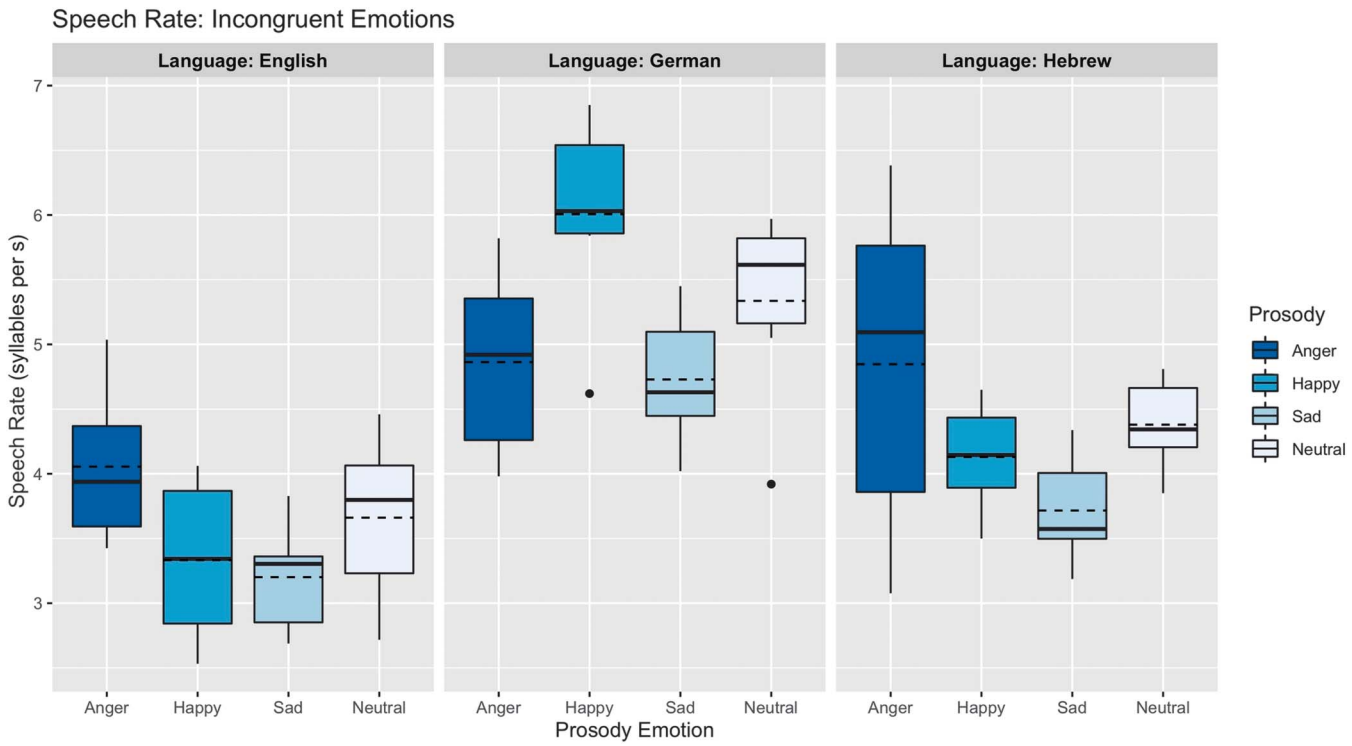
Results of the main analysis conducted for speech rate demonstrated significant main effects for prosody, $F(3, 69) = 7.572$ $p < .001$, $\eta_p^2 = 0.248$, and language, $F(2, 69) = 48.849$, $p < .001$, $\eta_p^2 = 0.586$, as well as a significant interaction between prosody and language, $F(6, 69) = 5.971$, $p < .001$, $\eta_p^2 = 0.34$. However, no significant main effect was found for congruency, $F(1, 69) = 0.126$, $p = .723$, nor did it lead to a significant interaction with language, $F(2, 69) = 2.606$, $p = .081$, with prosody, $F(2, 69) = 1.126$, $p = .33$, or a significant interaction of all three, $F(4, 69) = 1.371$, $p = .253$. In summary, it appears that speech rate was affected by the prosodic emotion and by the tested language. However, the semantic emotional content of the sentence, whether congruent or incongruent with the prosodic emotion, did not affect the speech rate.

Post hoc analysis of incongruent sentences demonstrated that speech rate was slowest for sadness, cross-linguistically, with significant contrasts for sadness versus anger (t.ratio = 3.086, $p < .05$) and sadness versus happiness (t.ratio = 2.655, $p < .05$). Within-language comparisons demonstrated a similar trend, with the sadness characterized by the slowest speech rate, for German (sadness vs. happiness, t.ratio = 3.223, $p = .01$) and Hebrew (sadness vs. anger, t.ratio = 2.852, $p < .05$). Figure 3 provides a visual depiction of this analysis.

Discussion

The current technical report compared the acoustic characteristics of three sets of spoken emotional sentences in English, German, and Hebrew, taken from the T-RES/

Figure 3. Box plots of speech rate values for incongruent sentences alone, across target prosodies and tested languages. The lower edge of the box represents the 25th quartile, the upper edge represents the 75th quartile (between them, the interquartile range; IQR), the solid line represents the median value, and the dotted line represents the mean value. The whiskers represent the highest and lowest values within 1.5 times the IQR. The solid dots represent outlying values (larger than 1.5 times the IQR).



iT-RES paradigms. Listeners' ratings, as collected in previous research, indicate that the T-RES prosodic sentences provide good representations of their respective emotional categories. The current report provides objective support for these subjective ratings. The acoustic measures of mean F0, F0 range, and speech rate have classically been used to describe and characterize different emotions (Bänziger & Scherer, 2005; Pell et al., 2009). Indeed, the acoustic characteristics can clearly differentiate the prosodic emotions one from another in each of the tested languages. The linguistic emotional content does not affect any of the tested acoustic prosodic features, indicating that the production of the prosody was not affected by the semantics. These findings provide a strong support for the validation of the tool as a gauge for emotional processing within a specific language. Finally, some common cross-linguistic trends were found. The emotion of neutrality was characterized by the lowest mean F0 and F0 range. Sadness was found to yield the slowest speech rate, and lower mean F0 than anger and happiness. These trends further support the validation of the T-RES for cross-linguistic comparisons and carry clinical implications for use of the tool, as described in the following sections.

Main Acoustic Findings

Mean F0

The use of mean F0 as a main acoustic measure of differentiating emotions is a common practice in the literature (e.g., Schmidt et al., 2016). The current findings of lower mean F0 for sadness than for anger and happiness emotions confirm previously reported findings in the literature. Sadness was consistently reported to be distinguished perceptually from other emotions, with a relatively low mean F0 (e.g., Pell et al., 2009). Anger, described as an "intense" emotion (Banse & Scherer, 1996), has been found to have a high mean F0 in comparison to other emotions (Drioli et al., 2003). Anger has also been demonstrated as easy to identify, cross-linguistically, in comparison to other emotions (Pell et al., 2009). It is not surprising to find that in the current study, the lowest mean F0 was noted for the neutral prosody, as in this condition, actresses were asked to imitate the formal tone of radio news broadcasters (see also Preti et al., 2016).

F0 Range

Prosodic emotions can be differentiated by F0 range (a significant main effect for prosody was found) across the tested languages to a similar extent (as no significant

interaction was found for prosody and language). Indeed, F0 range was reported in the literature to be an effective cue for prosodic discrimination (e.g., Banse & Scherer, 1996; Pell et al., 2009). It is noteworthy that F0 range, similar to mean F0, was found to be the smallest for neutrality, cross-linguistically. Indeed, adopting a flat affect diminishes the range of F0 used, as indicated previously in the literature (Ellgring & Scherer, 1996), whereas happiness and anger are characterized by high pitch variability (Banse & Scherer, 1996; Leitman et al., 2010). The current results of differentiation between emotional prosody for F0 range supports the use of this measure as a significant cue for affect perception (Schmidt et al., 2016).

Speech Rate

This study suggests clear within-language differences in speech rate between prosodic emotions, as well as a cross-linguistic trend suggesting slowest rates for sadness. This finding echoes the literature as sadness was consistently found to yield a slow speech rate across different languages (Thompson & Balkwill, 2006), whereas happiness and anger were generally found to have a much faster speech rate (Banse & Scherer, 1996; Leitman et al., 2010). On closer inspection, it appears that German sentences were characterized with fastest rates for happiness, whereas in English and Hebrew, anger yielded the highest rates. These findings echo the literature on German prosody. For example, Liu and Pell (2014) found the highest speech rate for the emotional prosody of happiness in German. The significant differences in speech rate between languages are also reflected in the literature (for a discussion, see Chu et al., 2021; Icht & Ben-David, 2014).

Validation

Taken together, the current findings confirm our hypotheses, and provide a strong support for the validation of the T-RES and iT-RES, in each of the tested languages, supporting behavioral (perceptual) rating data. First, emotional acoustic characteristics were found to be discriminable in terms of mean F0, F0 range, and speech rate. This indicates that listeners can easily identify the separate prosodic emotions based on commonly used acoustic features (Pell et al., 2009). Second, emotional prosodic features were not affected by the emotional semantics. In other words, the speakers produced highly similar prosodies for both congruent and incongruent spoken sentences (that convey the same/different emotion in the prosody and semantics). These findings demonstrate the lack of dependence of prosody and semantics in T-RES stimuli. This carries important clinical implications, as the T-RES gauges selective attention to one channel while inhibiting the emotional information in the other.

The current findings on cross-linguistic similarities in acoustic features further support the use of the tool with different linguistic populations (for discussions on cultural differences in vocal parameters, see Banse & Scherer, 1996, and Icht & Ben-David, 2014). Indeed, T-RES studies have indicated clear similarities in performance, notably, in all three languages a bias to process the prosodic emotion over the semantic one was indicated (Ben-David et al., 2016, 2019; Defren et al., 2018).

The results of the current report are of clinical significance, given the importance of emotional processing in daily communication and interactions (Icht, Zukerman et al., 2021) and the role of effective communication in well-being (Heinrich et al., 2016). The performance of clinical populations (e.g., individuals with ASD, individuals with hearing impairment) on the T-RES may shed light on this basic ability and may further guide intervention planning and the design of rehabilitation programs. In order to compare and generalize the results of this tool from one language to another, a cross linguistic validation as conducted in the current study is necessary. The acoustic analysis of the sets of stimuli facilitates adjusting the tool to populations who experience changes in hearing acuity (e.g., older adults; Ben-David et al., 2018).

Limitations and Future Directions

The main limitation of this report concerns the ability to generalize the acoustic findings to other tools, as sentences in the T-RES versions were recorded by professional actresses and not naturally occurring emotional speech. This limitation is in fact an advantage for the current validation of the tool, as it minimizes confounding factors.

The current report can serve to promote the cross-linguistic adaptation of other emotional speech tests (e.g., Florida Affect Battery [FAB]; Bowers et al., 1998; Diagnostic Analysis of Nonverbal Accuracy [DANVA]; Nowicki, 2000) and the adaptation of the T-RES to include additional languages. This may be especially revealing in languages that rely more heavily on prosodic information (e.g., tonal languages).

An additional limitation of the current report is the use of cross-linguistic comparisons with data from only a single, professional speaker for each of the three languages. The acoustic data obtained for each of the four emotional prosodies may potentially reflect speaker-specific instead of language-specific acoustic characteristics. This limitation was broadly addressed with comparisons of the current results to previous cross-linguistic data of emotional prosody patterns. Nonetheless, further research would greatly benefit from the inclusion of multiple speakers for a target language, to confirm language-specific acoustic characteristics.

Further studies may test the in-group processing advantage (Pell et al., 2009; Pell & Skorup, 2008)) with

listeners speaking the different languages (or a combination of two of the three languages), which may advance our understanding of the relative role of psychobiological and sociocultural factors in vocal emotion processing. Finally, the T-RES focuses on four basic emotions (anger, happiness, sadness, and neutrality). Further research may extend the tool to include more complex emotions (e.g., envy, boredom), as this distinction between simple and complex emotions was found to be of clinical importance (e.g., for individuals with ASDs, see Icht, Zukerman, et al., 2021).

Acknowledgments

This work was partially supported by the following grants awarded to the last author: the Israeli Science Foundation (Grant 861/18) and Ariel University Grant #RA2000000515.

References

Abu-Rabia, S. (2001). The role of vowels in reading semitic scripts: Data from Arabic and Hebrew. *Reading and Writing, 14*(1–2), 39–59. https://doi.org/10.1023/a:1008147606320

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication, 46*(3–4), 252–267. https://doi.org/10.1016/j.specom.2005.02.016

Ben-David, B. M., Ben-Itzchak, E., Zukerman, G., Yahav, G., & Icht, M. (2020). The perception of emotions in spoken language in undergraduates with high functioning autism spectrum disorder: A preserved social skill. *Journal of Autism and Developmental Disorders, 50*(3), 741–756. https://doi.org/10.1007/s10803-019-04297-2

Ben-David, B. M., Gal-Rosenblum, S., van Lieshout, P. H. H., & Shakuf, V. (2019). Age-related differences in the perception of emotion in spoken language: The relative roles of prosody and semantics. *Journal of Speech, Language, and Hearing Research, 62*(4S), 1188–1202. https://doi.org/10.1044/2018_JSLHR-H-ASCC7-18-0166

Ben-David, B. M., Malkin, G., & Erel, H. (2018). Ageism and neuropsychological tests. In L. Ayalon & C. Tesch-Römer (Eds.), *Contemporary perspectives on ageism. international perspectives on aging* (Vol. 19, pp. 279–297). Springer. https://doi.org/10.1007/978-3-319-73820-8_17

Ben-David, B. M., Mentzel, M., Icht, M., Gilad, M., Dor, Y. I., Ben-David, S., Carl, M., & Shakuf, V. (2020). Challenges and opportunities for telehealth assessment during COVID-19: IT-RES, adapting a remote version of the test for rating emotions in speech. *International Journal of Audiology, 60*(5), 319–321. https://doi.org/10.1080/14992027.2020.1833255

Ben-David, B. M., Multani, N., Shakuf, V., Rudzicz, F., & van Lieshout, P. H. H. (2016). Prosody and semantics are separate but not separable channels in the perception of emotional speech: Test for rating of emotions in speech. *Journal of Speech, Language, and Hearing Research, 59*(1), 72–89. https://doi.org/10.1044/2015_JSLHR-H-14-0323

Ben-David, B. M., Thayapararajah, A., & van Lieshout, P. H. H. (2013). A resource of validated digital audio recordings to assess identification of emotion in spoken language after a brain injury. *Brain Injury, 27*(2), 248–250. https://doi.org/10.3109/02699052.2012.740648

Ben-David, B. M., Van Lieshout, P. H. H., & Leszcz, T. (2011). A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain Injury, 25*(2), 206–220. https://doi.org/10.3109/02699052.2010.536197

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology, 9*(2), 78–84. https://doi.org/10.1027/1614-2241/a000057

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer (Version 6.1.07) (6.0.08)*.

Borden, G. J., & Harris, K. S. (1984). *Speech science primer: Physiology, acoustics, and perception of speech*. Lippincott Williams & Wilkins.

Bowers, D., Blonder, L. X., & Heilman, K. M. (1998). *Florida Affect Battery*. Center for Neuropsychological Studies.

Chu, S. Y., Lee, J., Barlow, S., Ben-David, B., Lim, K. X., & Foong, J. H. (2021). Oral-diadochokinetic rates among healthy Malaysian-Mandarin speakers: A cross linguistic comparison. *International Journal of Language & Communication Disorders, 23*(4), 419–429. https://doi.org/10.1080/17549507.2020.1808701

Defren, S., Wesseling, P. B. C., Allen, S., Shakuf, V., David, B. Ben, & Lachmann, T. (2018). *Emotional speech perception: A set of semantically validated German neutral and emotionally affective sentences*. Proceedings of the International Conference on Speech Prosody (pp. 714–718). https://doi.org/10.21437/SpeechProsody.2018-145

Drioli, C., Tisato, G., Cosi, P., & Tesser, F. (2003). Emotions and voice quality: Experiments with sinusoidal modeling. In *ISCA tutorial and research workshop on voice quality: Functions, analysis and synthesis, January* (pp. 127–132). ISCA Archive.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712–717. https://doi.org/10.1037/0022-3514.53.4.712

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion, 7*(1), 131–146. https://doi.org/10.1037/1528-3542.7.1.131

Ellgring, H., & Scherer, K. R. (1996). Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior, 20*(2), 83–110. https://doi.org/10.1007/BF02253071

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.

Heinrich, A., Gagné, J.-P., Viljanen, A., Levy, D. A., Ben-David, B. M., & Schneider, B. A. (2016). Effective communication as a fundamental aspect of active aging and well-being: Paying

attention to the challenges older adults face in noisy environments. *Social Inquiry Into Well-being, 2*(1), 51–68. https://doi.org/10.13165/SIIW-16-2-1-05

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Sage.

Hudepohl, M. B., Robins, D. L., King, T. Z., & Henrich, C. C. (2015). The role of emotion perception in adaptive functioning of people with autism spectrum disorders. *Autism, 19*(1), 107–112. https://doi.org/10.1177/1362361313512725

Hurley, D. S. (1992). Issues in teaching pragmatics, prosody, and non-verbal communication. *Applied Linguistics, 13*(3), 259–281. https://doi.org/10.1093/applin/13.3.259

Icht, M., & Ben-David, B. M. (2014). Oral-diadochokinesis rates across languages: English and Hebrew norms. *Journal of Communication Disorders, 48,* 27–37. https://doi.org/10.1016/j.jcomdis.2014.02.002

Icht, M., Wiznitser Ressis-tal, H., & Lotan, M. (2021). Can the vocal expression of intellectually disabled individuals be used as a pain indicator? Initial findings supporting a possible novice assessment method. *Frontiers in Psychology,* 2926. https://doi.org/10.3389/fpsyg.2021.655202

Icht, M., Zukerman, G., Ben-Itzchak, E., & Ben-David, B. M. (2021). Keep it simple: Identification of basic versus complex emotions in spoken language in individuals with autism spectrum disorder without intellectual disability: A meta-analysis study. *Autism Research, 14*(9), 1948–1964. https://doi.org/10.1002/aur.2551

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Komsta, L., & Novomestky, F. (2015). *Moments, cumulants, skewness, kurtosis and related tests*. R Package Version 0.14.

Leitman, D. I., Laukka, P., Juslin, P. N., Saccente, E., Butler, P., & Javitt, D. C. (2010). Getting the cue: Sensory contributions to auditory emotion recognition impairments in schizophrenia. *Schizophrenia Bulletin, 36*(3), 545–556. https://doi.org/10.1093/schbul/sbn115

Lenth, R. (2019). *emmeans: Estimated marginal means, aka least-squares means*. R Package Version 1.4.3.01. https://cran.r-project.org/package=emmeans

Leshem, R., Icht, M., Bentzur, R., & Ben-David, B. M. (2020). Processing of emotions in speech in forensic patients with schizophrenia: Impairments in identification, selective attention, and integration of speech channels. *Frontiers in Psychiatry, 11,* 1–13. https://doi.org/10.3389/fpsyt.2020.601763

Liu, P., & Pell, M. D. (2014). *Processing emotional prosody in mandarin Chinese: A cross-language comparison*. Proceedings of the International Conference on Speech Prosody (pp. 95–99). https://doi.org/10.21437/speechprosody.2014-7

Nowicki, S., Jr. (2000). *Manual for the receptive tests of the Diagnostic Analysis of Nonverbal Accuracy 2*. Emory University.

Oron, Y., Levy, O., Avivi-Reich, M., Goldfarb, A., Handzel, O., Shakuf, V., & Ben-David, B. M. (2020). Tinnitus affects the relative roles of semantics and prosody in the perception of emotions in spoken language. *International Journal of Audiology, 59*(3), 195–207. https://doi.org/10.1080/14992027.2019.1677952

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior, 33*(2), 107–120. https://doi.org/10.1007/s10919-008-0065-7

Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication, 50*(6), 519–530. https://doi.org/10.1016/j.specom.2008.03.006

Preti, E., Suttora, C., & Richetin, J. (2016). Can you hear what I feel? A validated prosodic set of angry, happy, and neutral Italian pseudowords. *Behavior Research Methods, 48*(1), 259–271. https://doi.org/10.3758/s13428-015-0570-7

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143–165. https://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R. (1989). Vocal correlates of emotion. In A. Manstead & H. Wagner (Eds.), *Handbook of psychophysiology: Emotion and social behavior* (pp. 165–197). Wiley. https://doi.org/10.1016/B978-0-12-558704-4.50015-3

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*(1), 76–92. https://doi.org/10.1177/0022022101032001009

Schmidt, J., Janse, E., & Scharenborg, O. (2016). Perception of emotion in conversational speech by younger and older listeners. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00781

Shakuf, V., Ben-David, B. M., Wegner, T., Wesseling, P. B. C., Allen, S., & Lachmann, T. (2022). *Hearing emotions across borders: German and Hebrew speakers process Hebrew prosody similarly* [Manuscript in preparation].

Shakuf, V., Gal-Rosenblum, S., & Ben-David, B. M. (2016). *The psychophysics of aging. In emotional speech, older adults attend to semantic, while younger adults to the prosody*. Fechner Day 2016: Proceedings of the 32nd Annual Meeting of the International Society for Psychophysics.

Taitelbaum-Swead, R., Icht, M., & Ben-David, B. M. (2022). *More than words: The relative roles of prosody and semantics in the perception of emotions in spoken language by postlingual cochlear implant recipients*. Ear and Hearing.

Thompson, W. F., & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica, 2000*(158), 407–424. https://doi.org/10.1515/SEM.2006.017