An Empirical Investigation on the Perceptual Similarity of Prosodic Language Types

Alina Gregori[1], *Frank Kügler*[1]

[1]Goethe University Frankfurt am Main, Germany

gregori@lingua.uni-frankfurt.de, kuegler@em.uni-frankfurt.de

Abstract

This study explores whether prosodic similarities of typological prosodic language types can be perceived by German native speakers. For this purpose, two online perception experiments were conducted containing twelve typologically and geographically diverse languages. Participants were asked to judge their prosodic similarity. Results showed that languages with mainly word-level prosodic properties were judged as similar to one another, while languages employing sentence-level prosodic properties were not clearly perceived as similar to their specific language type. Frequent confusions of Intonation and Phrase languages indicate a high perceptional similarity of languages belonging to these prosodic types. This leads to the assumption that the adopted distinction of prosodic properties is not completely represented in perception. Rather, additional prosodic factors influence the perception of the sentence prosody of languages.

Index Terms: prosodic typology, perception, prosodic properties, language similarity judgment, forced choice

1. Introduction

Prosodic typology groups languages according to word-level and sentence-level prosodic features (e.g. [1], [2], [3]). These features are often established through acoustic analyses of the prosodic inventory of a language. Based on these classifications, the present study aims at measuring the similarity of typologically closely related languages with regard to their sentence-level intonation patterns. For this purpose, two perception studies were conducted collecting similarity judgment data from languages belonging either to the same or to different prosodic language types. Inspired by the *Great Language Game* [4], our main question is whether prosodic language types are reflected in the perception of sentence prosody.

2. Background

2.1. Similarities between languages

Linguistic typology aims at identifying shared properties across languages to uncover the structural diversity of the world's languages. In a large-scale online survey, [4] investigated listeners' ability to identify a language (in an auditive forced-choice task). The authors analyzed whether the confusion patterns between languages correlate with linguistic similarities between languages based on typological classifications. As a result, [4] observed that participants more often confused two languages when they were more similar according to typological criteria such as similarity of sound inventories or shared lexical items. The study did not, however, take prosodic properties into account. On the basis of [4]'s findings, we hypothesize that languages that share prosodic properties belonging to the same prosodic language type would be confused as well. In other words, when investigating perceptual similarity, we expect languages of the same language type to be rated as similar. Our measure of similarity judgments departs from [4]'s forced choice guess paradigm.

2.2. Prosodic typology

There are different proposals on which prosodic properties may function as classification for prosodic language types (e.g. [1], [2], [3], [5], [6]). A classical distinction has been drawn on word-level prosodic properties to distinguish *tone* and *stress* systems [5], [7]. While acknowledging these two discrete language types, [5] argues for a continuous scale between the two types with "many intermediate word-prosodic systems which are not best seen as discrete types" [5, p. 226]. One such intermediate type is often referred to as 'pitch-accent language' such as Swedish or Japanese [8], who view 'accent languages' as one pole of a scale of accent and tone. However, [6] argues against such "a coherent prosodic type" ([6, p. 213]; cf. [9]) because languages may differ in which prosodic properties they employ.

Other proposals to prosodic typology include both word-level and sentence-level criteria [1, 2] distinguishing between different prominence and rhythmic properties at the word-level, and head-/edge-based and phrasing properties at the sentence level. As an extension, [2] proposed the property of *macro-rhythm* on the phrase as the crucial domain of distinction. It refers to the tonal pattern in utterance-medial position and is similar to a tonal density measure as proposed by [8, p. 35]. Concerning sentence-level prosody, [3] proposed to distinguish prominence and edge-based language types, introducing the new type of *phrase-languages*. These languages only have phrasal tones that vary in their position within the prosodic domain and are essentially not (or only loosely) associated with lexical stress [3, p. 226f].

2.3. Language types

For the classification of languages according to distinct language types we follow the proposal by [3] distinguishing between Intonation languages (I), Phrase languages (P), Pitch accent languages (A) and Tone languages (T). The languages chosen as representatives for each language type are listed in Table 1 below. Intonation languages mainly use sentence-level prosodic properties and no tonal word-level ones. This means that tones in such languages do not express lexical meaning but indicate prominence and head properties [3], [10]. Tone languages [5], [7], [11] are mainly characterized by word-level prosodic properties. Tone expresses lexical or grammatical meaning and is rarely used on higher prosodic domains. Pitch Accent languages use both sentence-level and word-level pro-

sodic properties [1], [2]. These languages distinguish words by lexical accents and use further tones to express sentence-level pragmatic meanings. The newly established language type, Phrase languages, associates tones at the phrasal level, expressing their demarcating function [3], [7]. The tone contour expresses prosodic structure.

2.4. Research question

Like [4], we assume a correlation between similar linguistic properties and languages of the same language type. We investigate whether listeners perceive similarities of languages based on sentence prosody. We hypothesize that, if participants perceive similarities in prosody, languages of the same language type will be judged as being more similar to one another than to languages of different language types.

3. Perception studies

Two perception experiments exercising different methods were conducted in order to examine the (dis-)similarity between languages with regard to their prosodic properties.

3.1. Speech materials

In our studies, we used three language samples per language type (12 in total, classified according to [3], see Table 1). Each language sample consisted of two sentences from the tale *The North Wind and the Sun*: "Then the Sun shone out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two." A native speaker of each language was recorded reading the complete tale.

Table 1: *Languages and their prosodic types*

I(ntonation)	P(hrase)	A(ccent)	T(one)
Portuguese	French	Swedish	Akan
Russian	Finnish	Serbian	Awing
Georgian	Urdu	Japanese	Igbo

3.2. Listeners

In experiment 1, 48 German native speakers (age range 15 - 64 years) participated. In experiment 2, 46 different German native speakers (age 16 - 69 years) participated. In a consent form, none of them indicated a hearing- or speech-impairment. Additionally, participants indicated their musicality on a scale of skilled (1) - not skilled (4) in a subjective estimation.

3.3. Procedure

Participants took part in an online experiment (created with the website SoSciSurvey [12]) listening to the language samples and judging their similarity based on "speech melody". Participants were allowed to listen to every stimulus multiple times.

Experiment 1 was a Single-Choice experiment where participants had to compare four languages to one target language and decide which of the four languages sounded the most similar (sentence prosody wise) to the target language. Each trial consisted of one language per language type, and a distinct language of one type was the target language. Every language in Table 1 was used as a target language once resulting in 12 language comparison trials. Across all participants, this makes 576 trials for Experiment 1.

Experiment 2 was a similarity judgment task. Each language was compared twice to other languages from the same type and two times to languages of the remaining types, resulting in 24 trials. The aim of experiment 2 is to provide a more direct comparison between two languages at a time. The participant was asked to judge on a scale (scroll bar) how similar the prosodic properties of the languages sounded to them (from 1 = very different, to 28 = very similar, the numbers were not visible to the participants). Overall, this results in 1.104 trials for Experiment 2.

3.4. Statistical measures

The statistical analysis was performed in R [13]. In the Single-Choice experiment, the independent variable was the language type of the target language, the dependent variable the language type of the chosen most similar language. A Confusion Matrix was created for illustration of the results, the "correct" answer counts are displayed in the diagonal cells from the upper left to lower right. Pearson's Chi-Square tests were performed as significance tests. Based on the residuals of the Chi-Square test, a correlation plot was plotted.

In the Similarity-Judgment experiment, the independent variable was the language type combination, i.e. the language types that were compared. The dependent variable was scalar (similarity judgment on a scale between 1 and 28). There were ten groups of language comparisons: four groups comparing languages of the same language types (138 data points each) and six groups comparing different language types (92 data points each). In order to minimize inter-subject variability, the data were normalized (for each trial, the value was divided by the participants' mean) resulting in a scale between 0 (not similar) and 5,42 (highest possible similarity). Mean values for the language type comparisons were illustrated in box plots. Statistical significance was calculated using a one-factorial ANOVA with TukeyHSD post-hoc tests.

4. Results

4.1. Single-Choice Experiment

In Table 2, similarity judgments between language types are displayed. Comparisons of the same language type show that only Tone languages had more matches (80; chance level: 36) with their language type than mismatches with others. Intonation, Phrase and Pitch Accent languages all show more mismatches of language types between target and chosen language, than matches. Considering the mismatches, it appears that Intonation languages were often confused with Phrase languages (58) and also with Pitch Accent languages (36), but not with Tone languages (11). Phrase languages were primarily confused with Intonation languages (72) and rarely confused with Tone languages (9). Pitch Accent languages were confused with all language types, but most often with Intonation languages (51). Tone languages show considerably low confusion with Intonation, Pitch Accent and Phrase languages.

Results from language type specific Chi-Square tests show that Tone languages are significantly perceived as similar (X-squared = 106.26, Df = 3, $p < 0.001$), indicating a coherent prosodic group in line with the hypothesis. The incorrect matches of Intonation and Phrase languages are significant as well (I: X-squared = 39.124, Df = 3, $p < 0.001$; P: X-squared = 20.082, Df = 3, $p < 0.001$). Matches and mismatches of Pitch accent languages return no significant result (X-squared = 1.9764, Df = 3, $p = 0.577$) suggesting that, contrary to the

hypothesis, this language type was not identified in perception.

Identical observations are displayed in the Correlation Plot (Figure 1), which shows correlations between the levels of the tested variables. The size of the dots indicates the strength of the respective correlation. The color shows the polarity of the correlation (blue = correlation, red = anticorrelation). Correlations can be observed in the comparison of Tone languages within their type and Intonation and Phrase languages against each other. All other comparisons are anticorrelations, thus the language types are judged as different to one another.

A striking observation from the Correlation Plot is that the results can be mirrored, cutting a line from the upper left corner to the lower right one. The correlations match their mirror image indicating that regardless of whether the target language or the chosen language is considered, the results are consistent.

Table 2: *Confusion Matrix Experiment 1*: *Rows display the independent variable, the columns the dependent one.*

	I	**P**	**A**	**T**	**total**
I	41	58	34	11	144
P	72	36	27	9	144
A	51	31	37	25	144
T	14	21	29	80	144
total	178	146	127	125	576

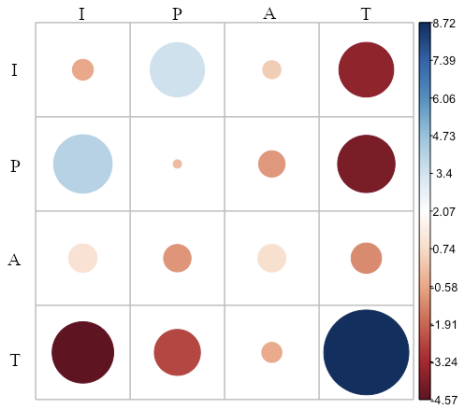

Figure 1: *Correlation plot based on residuals from Experimemt 1. Blue = correlation, Red = anticorrelation; dot size increases with strength of correlation.*

4.2. Similarity-Judgment Experiment

Looking at the language types in a more direct comparison within languages types (Table 3a), Tone languages are perceived as most similar (1,513), followed by Phrase languages (1,021) and Intonation languages (0,973). Pitch accent languages are perceived as the least similar (0,827). Comparing languages of different prosodic types (Table 3b), Intonation and Phrase languages received the highest similarity rating (1,429), higher than the ratings of their corresponding within-type pairings (Table 3a). Phrase and Tone languages are perceived as the least similar (0,492).

The box plots in Figure 2 not only illustrate these results, but also show the amount of variation within a language type comparison. The least variation is found in the comparisons T-T, I-P and P-T, all other language types show a bigger variation in their similarity judgment. The ANOVA showed that the observed similarity ratings are significant ($F = 30.8$, $Df = 9$, $p < 0.001$). Results of the post-hoc TukeyHSD tests confirmed that Tone languages among themselves and Intonation languages compared with Phrase languages were perceived as significantly more similar to each other than all other language type combinations. Phrase and Tone languages were rated as significantly less similar to all other language types. Thus, with regard to the hypothesis, the results suggest that listeners are able to perceive differences in prosodic properties, even if they are not always able to group the languages into the assumed types.

Table 3a & 3b: *Normalized data of Experiment 2; language type comparisons. Lowest possible similarity is 0, highest possible similarity is 5,42.*

I-I	**P-P**	**A-A**	**T-T**
0,973	1,021	0,827	1,513

I-P	**I-A**	**I-T**	**P-A**	**P-T**	**A-T**
1,429	0,947	0,777	1,053	0,492	0,787

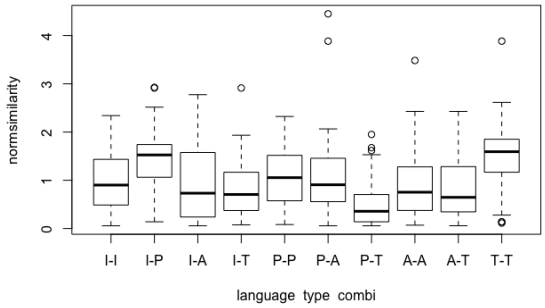

Figure 2: *Box plots Experiment 2; language type comparisons; for language types see Table 1.*

5. Discussion

With regard to the research question, both experiments reveal that prosodic language types are only partly distinguished in perception. The Single-Choice experiment showed that Tone languages were perceived as a distinct group, while Intonation, Phrase and Pitch accent languages are likely to be confused with other language types. In other words, while similarities in sentence prosody were perceived, the adopted language types were not correctly recognized, except for Tone languages. The Similarity-Judgment experiment provided a direct comparison between the languages and showed that the highest perceived similarity was between Tone languages (within their language type), followed by Intonation and Phrase languages, which were often confused. The language types perceived as the most distinct compared to one another were Phrase languages and Tone languages in direct comparison.

One reason for the strong disparity in perception between languages could be the prosodic background of the participants. In this study, all participants were native speakers of the Intonation language German. This might lead to a more differentiated judgment of Intonation languages opposed to less familiar languages (like Tone languages might be for native Germans). Thus, a natural expansion of the present study would be to test the same stimuli with native speakers of other language types to see whether the patterns of differentiation shift correspondingly. An indicator for this would be, for in-

stance, if native speakers of a Tone language perceive Intonation languages as very similar to one another, while rarely grouping Tone languages together, hence perceiving them as less similar.

While we strived to avoid familiarity with the languages, there is a high chance that it was not exclusively prosodic information that participants used for the similarity ratings. Since our stimuli were taken from recordings of reading out a tale, segmental information was present which is likely to influence perception. Although the instructions were to explicitly listen to the "speech melody", we cannot rule out the potential segmental influence. To minimize this influence, a follow up experiment would be reasonable, in which the stimuli are manipulated using low parse filters to rule out any natural language information and constrain the participants' attention to prosodic properties only.

The frequent confusion of Intonation and Phrase languages could also have a prosodic reason: While the tones in the two language types serve a different function in intonation, they both mark the same prosodic domains, namely phonological phrases [3]. In Intonation languages, pitch accents are used for prominence marking and they represent a head of a phonological phrase [14]. Phrase languages use tones to specify prosodic structure, i.e. they demarcate the domain of a phonological phrase [3]. They fulfill this purpose for speakers of the respective language, however for speakers that are not familiar with the languages, the tones might sound very similar. This could be a reason for the high rate of confusion of Intonation and Phrase languages, and deserves further investigation. In relation to that, the prosodic status of Georgian is still under discussion, but it was classified as an Intonation language in this article. Others, cf. [15], [16] have suggested that Georgian rather belongs to the Phrase language type. Between Intonation and Phrase languages in our data (Single-Choice Experiment), Georgian was rated most similar to a Phrase language in 72% of all trials, and only in 28% similar to an Intonation language. This might lead to the assumption that Georgian should be classified as a Phrase language. However, the mistaking of an Intonation language for a Phrase language was also common for the remaining Intonation languages. Thus, no final statement with regard to the language type of Georgian can be made based on the present data, which highlights the perceptional similarity of the two language types.

The observation that Pitch Accent languages are perceived as similar to all language types to a comparable extent is not surprising. Pitch accent languages have both sentence-level and word-level prosodic properties, making it such that they share features with all other language types. It was also suggested by e.g. [9], to not assume such a language type, as the languages assigned to this type differ so much in their prosodic properties, they can hardly comprise one group. Rather, they could be arranged on a scale from languages with sentence-level prosodic properties on one end, to languages with word-level prosodic properties on the other end. It would be possible to analyze the present data with regard to this suggestion to find out whether the similarity ratings of specific languages show a stronger similarity tendency towards any of the language types assumed here.

A reason for the missing ability to accurately group the examined languages into language types by perception could also be that similarity of languages of the same prosodic type are determined by other prosodic features. Dividing the languages and existing data according to other types based on factors like word-level prosodic properties [5], [7] or macrorhythm [2] might give further insight into which prosodic grouping is the most conceivable in perception.

While looking at language similarity, another factor shown to influence language identity rating [4] is geographical distance. Therefore, we compared the mean similarity values for the language pairs with the smallest geographical distance to the mean values of the relevant overall language type combinations (Table 4). No pattern in the difference in similarity ratings for languages with little geographical distance and the general language type comparisons is observable. Thus, it is assumed that geographical distance between the languages did not influence the results of the present study (contrary to [4]).

Table 4: *Geographical distance data*

Languages	**Similarity**	**Pros. types**	**Similarity**
Finnish - Swedish	1,024	P - A	1,053
Russian - Georgian	1,377	I - I	0,973
Awing - Igbo	1,395	T - T	1,513

Musicality increases accuracy in prosodic perception [17]. The subjective indication of musicality was divided into two groups: musically skilled (scale: 1 or 2) and not skilled (scale: 3 or 4). Musically skilled participants (about two thirds of all participants) showed that their similarity judgments on Tone languages were more precise with regard to prosodic grouping as well as scalar judgments than the judgments of musically not skilled participants (but musicality did not affect any of the other three language types).

6. Conclusion

This study investigated similarity judgments in typologically different languages. In short, we found that Tone languages were judged as sounding very similar to one another. The prosodic properties of this language type seem to be distinct to other languages, as matches of Tone languages with other types were not often observed. Intonation and Phrase languages were each rated as similar to the other language type more often than to their own language type indicating a high perceptual similarity of the prosodic properties of the two language types. Pitch Accent languages were rated as similar to all other languages to approximately the same extent, which shows that they are not easily assigned to one type.

To conclude, with regard to the hypothesis, our results show that, while differences in the sentence prosody of languages were perceived, the distinction of prosodic properties is partially, but not completely, represented in perception. Analyses with other prosodic groupings and further prosodic manipulation might show which prosodic factors ultimately influence the perception of sentence prosody of languages.

7. Acknowledgments

We thank the native speakers of the different languages for their recordings, and all participants for taking part in the studies. Financial support by Goethe University Frankfurt and by a DFG grant (KU 2323/4-1) is greatly acknowledged. Thanks to Beth Stoddard for English proofreading.

8. References

[1] S.-A. Jun, "Prosodic Typology," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed., Oxford: Oxford University Press, 2005, pp. 430–458.

[2] S.-A. Jun, Ed., *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 2014.

[3] C. Féry, *Intonation and Prosodic Structure*. Cambridge: Cambridge University Press, 2017.

[4] H. Skirgård, S. G. Roberts, and L. Yencken, "Why are some languages confused for others? Investigating data from the Great Language Game," *PloS one*, vol. 12, no. 4, 1-35, 2017, doi: 10.1371/journal.pone.0165934.L.

[5] M. Hyman, "Word-Prosodic Typology," *Phonology*, vol. 23, no. 2, pp. 225–257, 2006.

[6] L. M. Hyman, "How (not) to do phonological typology: the case of pitch-accent," *Language Sciences*, vol. 31, 2-3, pp. 213–238, 2009, doi: 10.1016/j.langsci.2008.12.007.

[7] H. van der Hulst, Ed., *Word Prosodic Systems in the Languages of Europe*. Berlin: Mouton De Gruyter, 1999.

[8] G. Bruce and O. Engstrand, "The phonetic profile of Swedish," *STUF-Sprachtypologie und Universalienforschung*, vol. 59, no. 1, pp. 12–35, 2006.

[9] C. Gussenhoven, *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press, 2004.

[10] D. R. Ladd, *Intonational Phonology*, 2nd ed. Cambridge: Cambridge University Press, 2008.

[11] L. M. Hyman, "Tone systems," in *Language Typology and Language Universals: An International Handbook*, M. Haspelmath, E. König, W. Oesterreicher, and W. Raible, Eds., Berlin: Walter de Gruyter, 2001, pp. 1367–1380.

[12] Leiner D. & S. Leiner, *SoSci der online Fragebogen.* Available: https://www.soscisurvey.de/de/index retrieved 30.04.2021

[13] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2013. Available: http://www.r-project.org/

[14] E. O. Selkirk, "Sentence prosody: Intonation, stress, and phrasing," in *The Handbook of Phonological Theory*, J. A. Goldsmith, Ed., Cambridge, MA: Blackwell, 1995, pp. 550–569.

[15] S. Skopeteas and C. Féry, "Focus and intonation in Georgian: constituent structure and prosodic realization," Ms, University of Göttingen and Goethe University Frankfurt, 2014.

[16] S. Skopeteas, C. Féry, and R. Asatiani, "Prosodic separation of postverbal material in Georgian: A corpus study on syntax-phonology interface," in *Information Structure in Lesser-described Languages: Studies in Prosody and Syntax*, E. Adamou, K. Haude, and M. Vanhove, Eds., Amsterdam: John Benjamins, 2018, pp. 17–50.

[17] N. Boll-Avetisyan, A. Bhatara, and B. Höhle, "Effects of musicality on the perception of rhythmic structure in speech," *Laboratory Phonology*, vol. 8, no. 1, pp. 1-16, 2017.