

Aesthetic perception of prosodic patterns as a factor in speech segmentation

Theresa Matzinger (theresa.matzinger@univie.ac.at)

University of Vienna, Department of Behavioral and Cognitive Biology, Althanstraße 14, 1090 Vienna, Austria
University of Vienna, Department of English, Spitalgasse 2, 1090 Vienna, Austria

Eva Specker (eva.specker@univie.ac.at)

University of Vienna, Department of Psychology, Liebiggasse 5, 1010 Vienna, Austria

Nikolaus Ritt (nikolaus.ritt@univie.ac.at)

University of Vienna, Department of English, Spitalgasse 2, 1090 Vienna, Austria

W. Tecumseh Fitch (tecumseh.fitch@univie.ac.at)

University of Vienna, Department of Behavioral and Cognitive Biology, Althanstraße 14, 1090 Vienna, Austria

Abstract

This study addresses the hypothesis that the aesthetic appeal of linguistic features may influence their learnability and in turn their stability in a language. Focusing on prosodic patterns, we investigated the crucial baseline assumption that linguistic features like stress affect aesthetic appeal. Listeners' liking, beauty and naturalness ratings of isochronous words and words with initially, medially or finally lengthened or shortened syllables revealed that, indeed, these patterns differed in their aesthetic appeal. Interestingly, the aesthetic appeal of prosodic patterns corresponded to their effectiveness for speech segmentation in other experiments, indicating a potential connection between aesthetics and language learning and opening up avenues for further research on the role of aesthetics in language acquisition and change.

Keywords: aesthetics; prosody; speech segmentation; artificial language learning; syllable duration

Introduction

All human cultures appreciate art and can perceive visual, verbal, or musical stimuli in terms of their aesthetic appeal (Nadal & Vartanian, 2019). While research on the explicit aesthetic appeal of linguistic features (such as Rastall, 2008) is still comparatively scarce, linguistic stimuli have often been investigated with regard to their processing fluency, emotional value, valence or arousal (Paulmann, Bleichner, & Kotz, 2013; Warriner, Kuperman, & Brysbaert, 2013) – features that are closely linked to aesthetic appeal (Leder, Ring, & Dressler, 2013; Reber, Schwarz, & Winkielman, 2004; Shibles, 1995). These connections make the more explicit investigation of the aesthetic appeal of linguistic features and its wider implications a worthwhile endeavor.

For example, among the aspects that produce the appeal of poetry, aesthetic preferences regarding prosodic patterns play a prominent role (Obermeier et al., 2016). If such preferences also apply in the perception of spontaneous everyday speech, they may affect language learning and, indirectly, also language change. A plausible hypothesis is that aesthetically appealing linguistic features are memorized more easily (Kousta, Vinson, & Vigliocco, 2009; Reber et al., 2004), used more frequently, and therefore transmitted more successfully

across speaker generations than features with less aesthetic appeal (cf. Smith & Kirby, 2008). Alternatively, features that violate aesthetic preferences may acquire an advantage in learning and transmission, because they may create negative arousal and thus be more easily noticed and remembered (e.g. Citron, Weekes, & Ferstl, 2014; Kuperman, Estes, Brysbaert, & Warriner, 2014). In either case, aesthetic appeal (or lack thereof) may constrain language learning, use, transmission, and change (e.g. sound changes or lexical borrowings; Rastall, 2008).

In this study, we investigate the crucial underlying assumption that there are in fact differences in people's aesthetic judgements of linguistic features, or patterns. To relate aesthetic judgements to language learning, we designed our study with respect to a widely investigated problem in language acquisition research, the speech segmentation problem (e.g. Saffran, Newport, & Aslin, 1996). This problem describes the challenge facing language learners to segment fluent speech into words. Several linguistic cues, including prosodic patterns in the speech stream, help learners to solve this problem (e.g. Matzinger, Ritt, & Fitch, 2021; Saffran et al., 1996; Tyler & Cutler, 2009). Therefore, we investigate listeners' aesthetic evaluation of different prosodic patterns, hypothesizing that the aesthetic appeal of prosodic patterns may affect the ease with which listeners can extract words from a continuous speech stream. If so, we could relate the aesthetic appeal of different prosodic patterns to speech segmentation, and draw more general conclusions about language learning and change. Crucially, when linking the aesthetic appeal of prosodic patterns to the role of these patterns in speech segmentation, we can only identify correlations, but not causalities.

Note that this study does not link aesthetic appeal and speech segmentation directly, i.e. within a single experiment, but rather provides a necessary baseline for such desirable future endeavors (see discussion) by explicitly testing in a word-rating experiment if prosodic patterns differ in their aesthetic appeal at all. When linking our experiment to speech segmentation, we refer to previous speech segmentation studies, most notably to one of our own studies

(Matzinger et al., 2021), which was similar to the present study in terms of stimulus and study design and therefore lends itself well for comparison.

We focus on the aesthetic perception of rhythmic patterns in trisyllabic words by native German speaking listeners. On the one hand, words may be regarded as most aesthetic if their syllables are isochronous, because isochrony has a facilitatory effect on auditory processing, and people have a general propensity for regular patterns (e.g. Poeppel & Assaneo, 2020; Ravignani & Madison, 2017). On the other hand, people also perceive irregular patterns as aesthetically appealing (e.g. Westphal-Fitch & Fitch, 2013). In that case, words with deviations from isochrony may be judged as more appealing than purely isochronous stimuli. The aesthetic perception of words with deviations from isochrony, especially on word-final syllables, is especially interesting with regard to the speech segmentation problem. We have shown in an artificial language learning experiment (Matzinger et al., 2021) that, when identifying trisyllabic words in a continuous speech stream, the performance of German speaking listeners improved when the final syllable of each word in the stream was lengthened. In contrast, their performance declined when the final syllable of each word was shortened. If aesthetic perception plays a role in speech segmentation, this predicts differences between the aesthetic appeal of words that have their final syllables lengthened and shortened, respectively. For example, if speech segmentation is facilitated by aesthetically appealing prosodic patterns, this predicts that words with lengthened final syllables should be rated highly and words with shortened final syllables should be rated lower in aesthetic appeal.

Besides durational deviations on word-final syllables, deviations on word-medial syllables are interesting for exploring the aesthetic appeal of prosodic patterns. Our previous speech segmentation experiment (Matzinger et al., 2021) showed that native German speaking participants preferably segmented words with medially shortened syllables from a continuous speech stream. If speech segmentation is facilitated by a high aesthetic appeal, participants should therefore perceive words with shortened medial syllables as aesthetically appealing. On the other hand, durational changes might be the result of language-specific stress patterns, in which, usually, stressed syllables are lengthened, and unstressed syllables shortened or reduced (Ordin, Polyanskaya, Laka, & Nespors, 2017; Tyler & Cutler, 2009). Due to an exposure effect (Sluckin, Hargreaves, & Colman, 1983; Zajonc, 1968), listeners may find durational variations that match the typical stress patterns of their native language more aesthetically appealing than other durational variations (but there are also alternative theories stating that moderately frequent or novel, i.e. infrequent items may be perceived as most aesthetically appealing; Martindale et al., 1988). Our participants were native speakers of German, and most German trisyllabic words are stressed on their word-medial syllable (Domahs, Plag, & Carroll, 2014; Ernestus & Neijt, 2008). Therefore, if participants prefer durational patterns that match the typical stress patterns of their native

language, they should rank words with lengthened medial syllables as aesthetically appealing.

To complement our investigations on words with word-final and word-medial durational modifications, we also tested the aesthetic appeal of words with word-initial durational modifications. However, we did not have clear predictions for the aesthetic appeal of words with initially modified syllables based on our previous speech segmentation experiment or prosodic features of German.

To summarize, our study investigated how listeners judged the aesthetic appeal of trisyllabic words where word-initial, word-medial or word-final syllables were lengthened or shortened, compared to an isochronous baseline. We measured aesthetic appeal by collecting participants' ratings of these different prosodic patterns. They ranked each acoustic stimulus on its 'liking', its 'beauty' and its 'naturalness'. Although closely related, these concepts are different manifestations of aesthetic appeal (Conway & Rehding, 2013). 'Liking' refers to purely sensual pleasure and is often regarded as an explicit evaluative judgement, whereas 'beauty' requires higher executive functions and includes more emotional involvement (Armstrong & Detweiler-Bedell, 2008; Briemann & Pelli, 2017). 'Naturalness' serves as a control category to test if judgements of likability and beauty might be influenced by how natural participants find the stimuli.

Ratings of all three measures of aesthetic appeal might be influenced by the occurrence frequencies of the respective prosodic patterns in the participants' native language, German (Sluckin et al., 1983). However, 'naturalness' is expected to be most strongly influenced by occurrence frequency, with most frequent prosodic patterns being judged as most natural. Therefore, naturalness ratings might help to determine if liking and beauty are influenced by the occurrence frequency of the respective prosodic patterns.

Methods

Experimental Conditions and Procedure

We tested if participants perceived trisyllabic pseudo-words that either had isochronous syllables (isochrony condition) or had one of their syllables lengthened (lengthening condition) or shortened (shortening condition) as differing in their aesthetic appeal. We conducted three experiments: in the first experiment, lengthening and shortening happened on word-initial syllables, in the second experiment on word-medial syllables, and in the third experiment on word-final syllables. Otherwise, the three experiments were identical in their experimental procedure. Thus, durational condition (isochrony, lengthening, and shortening) was a within-subjects variable and modification position (word-initial, word-medial, and word-final) a between-subjects variable.

For each of 20 pseudo-word stimuli (Tab. 1), each participant rated its liking, beauty and naturalness. Ratings were blocked on liking and beauty: liking and beauty were counterbalanced, and naturalness was always the last block, i.e., participants first rated all stimuli on liking before rating

the same stimuli on beauty (or vice versa), and finally gave naturalness ratings. All participants completed naturalness ratings last because naturalness ratings might be heavily biased by the stage of occurrence in the experiment (e.g. early in the experiment, artificial pseudo-words might be regarded as more unnatural than later in the experiment when participants have already been exposed to many similar pseudo-words) and because the naturalness ratings served as a control category to test if participants' judgements of liking and beauty were influenced by their naturalness. Participants ranked each stimulus on a scale from 1 to 7, with 1 being the least and 7 being the most likable, beautiful, or natural.

In total, in each experiment each participant ranked each word nine times, namely in three durational conditions (isochrony, lengthening and shortening) and for three manifestations of aesthetic appeal (liking, beauty and naturalness). Modification positions (either word-initial, word-medial, or word-final position) varied across experiments, meaning that each participant only heard words modified in one position.

Table 1: Artificial words used in the study. Hyphens indicate syllable boundaries.

ba-pe-di	fa-ro-vu	ke-ta-fi	me-ko-ru	ri-fe-tu
bo-re-fu	fo-pu-ve	ku-te-so	mu-lo-se	si-go-va
da-ni-mo	ga-su-de	le-vi-po	ne-bu-pa	ti-nu-ge
do-mi-ka	gu-sa-ki	lu-bi-na	pi-ma-to	vo-la-gi

Participants and Setting

In total, 180 monolingual native German speaking participants with English as a second language took part in all 3 experiments: 60 people each participated in the word-initial modification experiment (43 female, mean age: $20.6 \pm SD 1.98$; 2 participants were excluded because of technical issues during data collection), the word-medial modification experiment (43 female, mean age: $20.5 \pm SD 1.67$) and the word-final modification experiment (39 female, mean age: $21.4 \pm SD 3.23$). Testing lasted about 30 minutes per participant and participants were rewarded with study credits for their participation.

Participants were tested in a laboratory setting that was free of background noise. Testing was administered via desktop computers and the experimental interface that was used to present the stimuli and collect ratings was created in OpenSesame (version 3.1.9; Mathôt, Schreij, & Theeuwes, 2012). Participants gave their ratings by pressing keys on a computer keyboard with their preferred hand. Stimuli were presented binaurally over Sennheiser HD 300 PRO headphones and had the same amplitude for all participants.

Stimuli

Creation of the Artificial Words The artificial language used in these experiments consisted of 20 words (Tab. 1), which all consisted of three consonant-vowel syllables. To

avoid priming between artificial words and possible effects of different intrinsic aesthetic values of different consonant-vowel combinations, each consonant-vowel combination, i.e. each syllable, only occurred once in our set of words.

Creation of the Sound Signals To create the actual sound signals for the experiments, each syllable was recorded individually by a female native speaker of German. The acoustic parameters of the syllables were normalized in Praat (version 6.0.36; Boersma & Weenik, 2017) so that each syllable had a duration of 400 ms and a fundamental frequency (f_0) of 210 Hz. In addition, syllable amplitude was normalized by scaling the amplitude of each syllable so that its absolute peak amplitude was 0.99. To create the stimuli for the lengthening and shortening conditions, the respective syllables were modified in Praat so that their duration was 150% or 50% of their original duration, making the shortened syllables 200 ms long and the lengthened syllables 600 ms long. Finally, the normalized and modified syllables were concatenated to form artificial words (Tab. 1) using custom code in Python 3.6.3.

Results

Descriptive Statistics: Mean Liking, Beauty and Naturalness Ratings

The mean liking, beauty and naturalness ratings with their confidence intervals (Fig. 1) show that, overall, there were differences (indicated by non-overlapping confidence intervals) in aesthetic appeal between different durational modifications at different positions within trisyllabic words. In general, words with shortened syllables were rated as less aesthetically appealing than words with isochronous or lengthened syllables, and this effect manifested most prominently when shortening happened word-finally (for more details see the model results below). However, overall, differences were small, with mean ratings ranging from 2.83 ($SD = 1.50$; mean liking of word-finally shortened words) to 3.83 ($SD = 1.76$; mean naturalness of isochronous syllables in the experiment on word-initial modifications) on a seven-point rating scale (Fig. 1).

In general, the mean ratings of liking and beauty were similar, whereas naturalness ratings were slightly higher (see non-overlapping confidence intervals in Fig. 1 and results of the Cumulative Link Mixed Models below), indicating that participants' sense for naturalness was more liberal than their sense for liking and beauty.

To investigate possible correlations between the ratings on liking, beauty and naturalness, we calculated Spearman's rank correlation coefficients (Spearman, 1910), using the `cor.test` function in R (version 3.6.0; R Development Core Team, 2018). This revealed that there was a moderate positive correlation (Cohen, 1992) between liking and beauty ratings ($r_s(10,678) = 0.50, p < 0.001$), liking and naturalness ratings ($r_s(10,678) = 0.46, p < 0.001$) and beauty and naturalness ratings ($r_s(10,678) = 0.46, p < 0.001$).

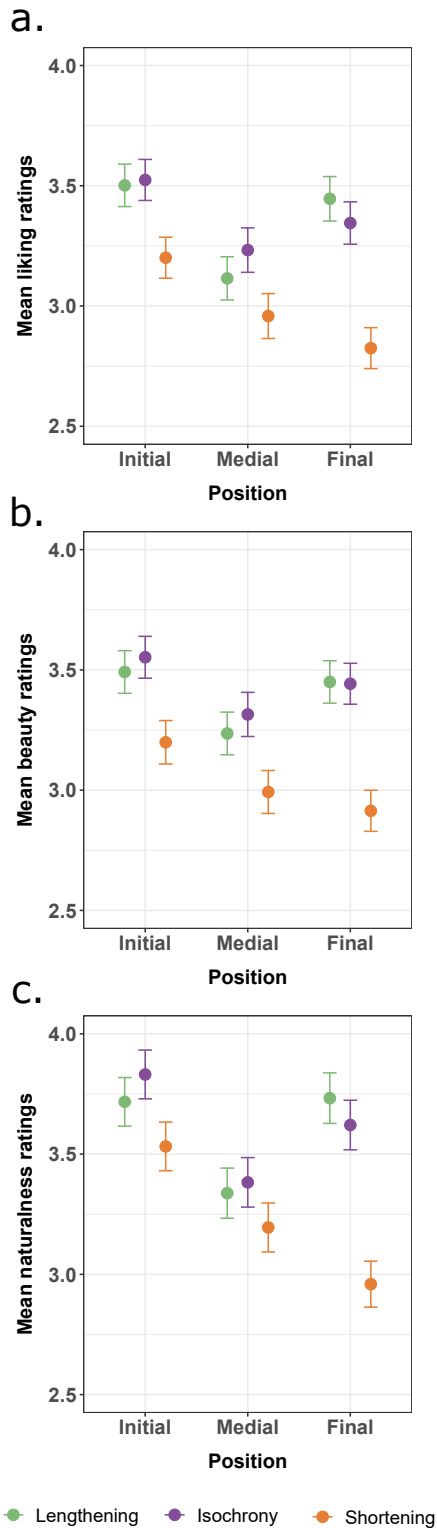


Figure 1. Mean a) liking, b) beauty, and c) naturalness ratings of words with isochronous, lengthened or shortened syllables in word-initial, word-medial and word-final position. Participants rated the stimuli's liking, beauty and naturalness on a scale from 1 (least likable, beautiful and natural) to 7 (most likable, beautiful and natural). Error bars denote 95% confidence intervals.

Cumulative Link Mixed Model

Analysis To test if the aesthetic perception of the artificial pseudo-words was influenced by *durational condition* and *modification position*, i.e. whether different durational modifications led to different ratings when they occurred in different positions, and whether this differed for *liking*, *beauty* and *naturalness*, we applied a Cumulative Link Mixed Model (Christensen, 2018, 2019a; Christensen & Brockhoff, 2013), using Laplace approximation (Joe, 2008; Pinheiro & Bates, 1995). One advantage of CLMMs over standard linear mixed models is that they reflect the ordinal nature of the rating scale data (Christensen & Brockhoff, 2013).

In this model, durational condition and modification position, as well as their interaction were included as fixed effects. The type of aesthetic appeal (i.e. liking, beauty and naturalness) was included as an additional fixed effect. We also entered random intercepts of participant and word into the model. To avoid inflated type I error rates, each model included a random slope (Barr, Levy, Scheepers, & Tily, 2013) of durational condition within participant. The sample size for the model was 32,040 data points (178 individuals tested for 3 types of aesthetic appeal on 3 durational conditions with 20 words each). We used the durational condition of isochrony, the word-medial modification position and the aesthetic appeal of naturalness as reference levels in the model.

The model was fitted in R (version 3.6.0; R Development Core Team, 2018), using the function `clmm` of the R-package `ordinal` (version 2019.12.10; Christensen, 2019b).

We used a likelihood ratio test to test the overall significance of the full model as compared to a null model comprising only the random effects (R function `anova`; Dobson, 2002). P-values for the effects of individual predictors are based on likelihood ratio tests that compare the full model with a reduced model lacking the fixed effects one at a time (R function `drop1`; Barr et al., 2013).

Results Overall, the full model (Tab. 2) was significantly different from the null model, indicating an effect of durational condition, its potential interaction with modification position or of the type of aesthetic appeal on how participants rated the stimuli (likelihood ratio test: $\chi^2 = 212.0$, $df = 10$, $p < 0.001$). More specifically, we found that the durational condition influenced the ratings of stimuli (likelihood ratio test: $\chi^2 = 72.2$, $df = 2$, $p < 0.001$): in general, lengthening and shortening of syllables within a word had a negative effect on the ratings compared to isochrony (Tab. 2). Additionally, the interaction between durational condition and modification position had a significant effect on the ratings of stimuli (likelihood ratio test: $\chi^2 = 16.0$, $df = 4$, $p = 0.003$). This indicates that the ratings of durational variations were influenced by the syllable position at which the durational variations occurred. Especially, word-finally lengthened syllables had a positive effect on the ratings (despite the overall negative effect of lengthening), whereas word-finally shortened syllables had an additional negative effect on the ratings (Tab. 2, Fig. 1). Also, there was a

significant effect of the type of aesthetic appeal on the ratings (likelihood ratio test: $\chi^2 = 112.0$, $df = 2$, $p < 0.001$): participants provided lower ratings for liking and beauty as compared to naturalness (Tab. 2). Overall, there was no effect of modification position on the ratings (likelihood ratio test: $\chi^2 = 5.1$, $df = 2$, $p = 0.077$), but participants tested on initial modifications provided slightly higher rankings than participants tested on medial modifications (Tab. 2).

Table 2. Results of the Cumulative Link Mixed Model exploring the effects of duration, position, their interaction and type of aesthetic appeal on the ratings of participants. The table reports estimated model coefficients (Est.), standard errors (SE), z-values (z) and p-values (p) of the fixed effects.

Coefficients	Est.	SE	z	p
appealLiking	-0.26	0.03	-10.28	< 0.001
appealBeauty	-0.19	0.03	-7.56	< 0.001
durationLonger	-0.11	0.05	-2.23	0.026
durationShorter	-0.38	0.10	-3.68	< 0.001
positionInitial	0.48	0.22	2.18	0.029
positionFinal	0.25	0.22	1.13	0.257
durationLonger: positionInitial	0.03	0.07	0.41	0.684
durationShorter: positionInitial	-0.07	0.15	-0.44	0.659
durationLonger: positionFinal	0.20	0.07	2.92	0.004
durationShorter: positionFinal	-0.44	0.15	-3.04	0.002

Discussion

To summarize, in general, words with one syllable shortened had a lower aesthetic appeal than isochronous words. In contrast, words with one of their syllables lengthened had a lower aesthetic appeal than isochronous words only when lengthening occurred word-initially or word-medially, but not when it occurred word-finally (see results of the CLMM). Overall, this finding suggests that humans may have a more general preference for regular and isochronous patterns in words (e.g. Ravnani & Madison, 2017).

One explanation for why participants rated words with shortened syllables as less appealing could be that shortened syllables indicate a fast speech rate. Fast speech is often associated with low-prestige and casual situations, or with nervousness (Demenko & Jastrzębska, 2012), and listeners may have disliked stimuli with shortened syllables for that reason. This would point towards a socially learned connection between prosodic patterns and aesthetic appeal.

The low aesthetic appeal of words with shortened syllables manifested most prominently when shortening happened on word-final syllables (non-overlapping confidence intervals in Fig. 1). This is in line with our previous speech segmentation experiment (Matzinger et al., 2021), in which final syllable shortening hindered the segmentation of words from a continuous speech stream. Together, these findings indicate

that a low aesthetic appeal of prosodic patterns may correlate with a disadvantage of these patterns for speech segmentation, and potentially for language learning in a wider sense. The causalities behind these correlations are however unclear: the correlations may indicate either that listeners do not use aesthetically unpleasant prosodic patterns for speech segmentation, or that patterns that are not used for speech segmentation for any reason, are not perceived as aesthetically appealing. Most probably, both directions of this relationship interact, or additional underlying factors such as acoustic salience, speech rate, context, occurrence frequencies, neural oscillations, ease of processing or memory influence them both (Forster, Leder, & Ansorge, 2013; Morrill et al., 2015; Obermeier et al., 2016; Palmer & Mattys, 2016; Poeppel & Assaneo, 2020; Reber et al., 2004).

Words with finally lengthened syllables, which boosted speech segmentation in our previous study (Matzinger et al., 2021), did not differ in their aesthetic appeal from isochronous words (overlapping confidence intervals in Fig. 1). This indicates that the positive effect of word-final lengthening for speech segmentation may not be related to its aesthetic appeal, but rather to other underlying factors (as discussed in the previous paragraph). The differences in aesthetic appeal of words with finally lengthened and finally shortened syllables suggest that prosodic patterns that are aesthetically appealing do not support speech segmentation, but patterns that are unappealing can hinder it.

In our previous speech segmentation experiment (Matzinger et al., 2021), shortened syllables were preferably segmented as occurring word-medially. This led to the prediction for this study that, if aesthetic appeal facilitates speech segmentation, shortened medial syllables should be ranked as most aesthetically appealing. This prediction, however, was not borne out because – like in word-initial and word-final position – also in word-medial position, shortened syllables were perceived as less likable and beautiful than lengthened and isochronous syllables (non-overlapping confidence intervals in Fig. 1). Still, the negative effect of medial shortening on likability and beauty was smaller than the negative effect for initial and final shortening, and there was no negative effect of medial shortening on naturalness (Fig. 1). This may reflect that initial and final syllables are particularly important for speech processing and segmentation, while word-medial syllables are less relevant (Tyler & Cutler, 2009; Wedel, Ussishkin, & King, 2019). Although in this case, evidence is undoubtedly weak, indirect and not as straight-forward as we had predicted, it seems plausible that this finding reflects a link between aesthetic appeal and speech segmentation performance.

Interestingly, we did not find naturalness ratings of prosodic patterns to be influenced by the occurrence frequencies of stress patterns in the participants' native language. Since most German trisyllables are stressed on the word-medial syllable, and stress usually correlates with lengthening (Ordin et al., 2017), we would have expected words with lengthened word-medial syllables to be most natural, and words with shortened word-medial syllables

least natural. Instead, we found that the naturalness of medially lengthened and medially shortened words was comparable. Interestingly, there was a large difference in naturalness ratings of words with finally lengthened and finally shortened syllables: finally lengthened words were rated as much more natural. This finding may be explained by the occurrence frequencies of prosodic patterns in natural languages. Cross-linguistically, final syllables are lengthened for multiple reasons, for example to indicate boundaries (Fletcher, 2010). This boundary-related lengthening mostly happens phrase-finally, and not word-finally, but since in our design, each word essentially equaled a phrase, participants may have transferred the high naturalness of phrase-final lengthening to our stimuli.

Methodologically, we decided against including other prosodic cues such as variations in f_0 or intensity in our stimuli to be able to clearly attribute the differences in aesthetic appeal to the durational modifications. Undoubtedly, leaving out all other prosodic cues comes at the cost of the stimuli being slightly unnatural from the outset. For example, in natural languages, stress has multiple correlates, and participants in our experiments may not have interpreted syllables that only had a longer duration, but not a higher intensity or a higher f_0 as being stressed (cf. Gordon & Roettger, 2017). Also, in natural languages, pitch and intensity typically decrease phrase-finally, which we did not model into our stimuli (Vaissière, 1983). Future experiments evaluating the aesthetic appeal of prosodic patterns could combine different voice modulatory cues in their stimuli, which would resemble natural speech more closely and provide a more nuanced picture. Also, to make the task more natural from the outset, target words could be embedded in a bigger context (e.g. in an utterance) in future experiments.

Liking, beauty and naturalness ratings were positively correlated, indicating that participants perceive these three concepts to be highly related. This suggests that in future studies, it might be sufficient to use a single measure of aesthetic appeal, and that data across different dimensions of aesthetic appeal may be combined.

The fact that naturalness ratings were slightly higher than liking and beauty ratings suggests that naturalness ratings were provided less conservatively than liking and beauty ratings. Participants possibly did not find our stimuli particularly appealing in general (e.g. for reasons discussed in the previous paragraph). This dislike may be less reflected in the naturalness ratings because naturalness may be influenced less by explicit aesthetic judgements than liking and beauty. An alternative explanation for this finding could be that in our study design, naturalness ratings were always provided last, and participants may have provided higher ratings after a longer exposure to the stimuli (Sluckin 1983).

Conclusion

To conclude, we showed that different prosodic patterns differed in their likability, beauty and naturalness. This finding is an important baseline for further investigating the potential relationship between aesthetic appeal, prosodic

patterns and cognitive factors such as learnability, ease of processing or memory. In connection with the effectiveness of different prosodic cues for speech segmentation, our findings on aesthetic appeal make it plausible that such a relationship exists. Since our study provides a crucial prerequisite for the study of aesthetic perception as a potential bias in language learning and language change, it opens up avenues for future research. A deeper insight into the role of aesthetic perception of linguistic features as a potential bias in language learning and language change could for example be gained by studies that test this relationship in a more direct and explicit way (e.g. iterated learning experiments).

References

- Armstrong, T., & Detweiler-Bedell, B. (2008). Beauty as an Emotion: The Exhilarating Prospect of Mastering a Challenging World. *Review of General Psychology*, *12*(4), 305–329.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.
- Boersma, P., & Weenik, D. (2017). Praat: doing phonetics by computer.
- Briellmann, A. A., & Pelli, D. G. (2017). Beauty Requires Thought. *Current Biology*, *27*(10), 1506-1513.e3.
- Christensen, R. H. B. (2018). Cumulative Link Models for Ordinal Regression with the R Package ordinal. *Journal of Statistical Software*, 1–40.
- Christensen, R. H. B. (2019a). A Tutorial on fitting Cumulative Link Models with the ordinal Package.
- Christensen, R. H. B. (2019b). ordinal - Regression Models for Ordinal Data.
- Christensen, R. H. B., & Brockhoff, P. (2013). Analysis of sensory ratings data with cumulative link models. *Journal de La Société Française de Statistique & Revue de Statistique Appliquée*, *154*(3), 58–79.
- Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2014). Arousal and emotional valence interact in written word recognition. *Language, Cognition and Neuroscience*, *29*(10), 1257–1267.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, *112*(1), 155–159.
- Conway, B. R., & Rehding, A. (2013). Neuroaesthetics and the Trouble with Beauty. *PLoS Biology*, *11*(3), 1–5.
- Demenko, G., & Jastrzębska, M. (2012). Analysis of natural speech under stress. *Acta Physica Polonica A*, *121*(1 A), 92–95.
- Dobson, A. J. (2002). *An introduction to generalized linear models*. Boca Raton: Chapman & Hall.
- Domahs, U., Plag, I., & Carroll, R. (2014). Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited. *The Journal of Comparative Germanic Linguistics*, *17*(1), 59–96.
- Ernestus, M., & Neijt, A. (2008). Word length and the location of primary word stress in Dutch, German, and English. *Linguistics*, *46*(3), 507–540.

- Fletcher, J. (2010). The Prosody of Speech: Timing and Rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (2nd ed., pp. 523–602). Hoboken: Wiley-Blackwell.
- Forster, M., Leder, H., & Ansorge, U. (2013). It felt fluent, and i liked it: Subjective feeling of fluency rather than objective fluency determines liking. *Emotion, 13*(2), 280–289.
- Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard, 3*(1), 1–11.
- Joe, H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis, 52*(12), 5066–5074.
- Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112*(3), 473–481.
- Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: valence and arousal affect word recognition. *Journal of Experimental Psychology. General, 143*(3), 1065–1081.
- Leder, H., Ring, A., & Dressler, S. G. (2013). See me, feel me! Aesthetic evaluations of art portraits. *Psychology of Aesthetics, Creativity, and the Arts, 7*(4), 358–369.
- Martindale, C., Moore, K., & West, A. (1988). Relationship of Preference Judgments to Typicality, Novelty, and Mere Exposure. *Empirical Studies of the Arts, 6*, 79–96.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324.
- Matzinger, T., Ritt, N., & Fitch, W. T. (2021). The Influence of Different Prosodic Cues on Word Segmentation. *Frontiers in Psychology, 12*.
- Morrill, T. H., McAuley, J. D., Dille, L. C., Zdziarska, P. A., Jones, K. B., & Sanders, L. D. (2015). Distal prosody affects learning of novel words in an artificial language. *Psychonomic Bulletin and Review, 22*(3), 815–823.
- Nadal, M., & Vartanian, O. (Eds.). (2019). *The Oxford Handbook of Empirical Aesthetics*. Oxford: Oxford University Press.
- Obermeier, C., Kotz, S. A., Jessen, S., Raettig, T., von Koppenfels, M., & Menninghaus, W. (2016). Aesthetic appreciation of poetry correlates with ease of processing in event-related potentials. *Cognitive, Affective and Behavioral Neuroscience, 16*, 362–373.
- Ordin, M., Polyanskaya, L., Laka, I., & Nespore, M. (2017). Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory & Cognition, 45*, 863–876.
- Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *Quarterly Journal of Experimental Psychology, 69*(12), 2390–2401.
- Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology, 4*, 1–10.
- Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics, 4*(1), 12–35.
- Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience, 21*(6), 322–334.
- R Development Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Rastall, P. (2008). Aesthetic responses and the “cloudiness” of language: Is there an aesthetic function of language? *La Linguistique, 44*, 103–132.
- Ravignani, A., & Madison, G. (2017). The paradox of isochrony in the evolution of human rhythm. *Frontiers in Psychology, 8*, 1–13.
- Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver’s processing experience? *Personality and Social Psychology Review, 8*(4), 364–382.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language, 35*, 606–621.
- Shibles, W. (1995). *Emotion in aesthetics*. Dordrecht: Springer Science + Business Media.
- Sluckin, W., Hargreaves, D. J., & Colman, A. M. (1983). Novelty and human aesthetic preferences. In J. Archer & L. I. A. Birke (Eds.), *Exploration in animals and humans* (pp. 245–269). Wokingham: Van Nostrand Reinhold.
- Smith, K., & Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 363*(1509), 3591–3603.
- Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America, 126*(1), 367–376.
- Vaissière, J. (1983). Language-Independent Prosodic Features. In A. Cutler & D. R. Ladd (Eds.), *Springer Series in Language and Communication 14: Prosody: Models and Measurements* (pp. 53–66). Hamburg: Springer.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207.
- Wedel, A., Ussishkin, A., & King, A. (2019). Incremental word processing influences the evolution of phonotactic patterns. *Folia Linguistica, 40*, 231–248.
- Westphal-Fitch, G., & Fitch, W. T. (2013). Spatial Analysis of “Crazy Quilts”, a Class of Potentially Random Aesthetic Artefacts. *PLoS ONE, 8*(9).
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1–27.