Iconicity versus indexicality in language attitudes: The case of /x/

Simon David Stein
Heinrich Heine University Düsseldorf, Germany,
Faculty of Arts and Humanities, Department of English and American Studies

Abstract

French sounds beautiful and romantic, German sounds harsh and aggressive—or so many people think. But why? What are the cognitive mechanisms behind such language attitudes? Traditionally, attitudes have been explained by indexicality—evaluative meaning emerges because linguistic features point to social groups. However, research on iconicity suggests that such meaning can emerge due to properties of sounds themselves. Using newly constructed language stimuli, this study investigated the role these two mechanisms play for the perception of the voiceless velar fricative /x/. 501 listeners rated stimuli with and without /x/ on several semantic scales. Results indicate that /x/ has little direct effect, but listeners with prior exposure to /x/ rated it more negatively, which highlights the importance of sociocultural associations. Overall, the findings indicate that language attitudes are strongly shaped by indexical factors such as exposure and familiarity, and that iconicity is modulated by indexicality.

Keywords: language attitudes, iconicity, indexicality, sociolinguistics, phonaesthetics

1. Introduction

Speech sounds do not merely distinguish meaning, they have meaning. They carry social meaning that makes listeners have attitudes towards the speech signal, and towards the person producing it. These attitudes can manifest as ratings of perceived character traits like a speaker's friendliness, intelligence, or trustworthiness, as aesthetic judgments about beauty, shape, size, texture, and as simple evaluations of general likeability or pleasantness (see, e.g., Coupland & Bishop, 2007; Winter et al., 2022; Anikin et al., 2023). Consequently, one language (e.g., French) may stereotypically be perceived as sounding more "pleasant" than another language (e.g., German). Or, one particular variety of a language (e.g., Southern Irish English) may stereotypically be perceived as more pleasant-sounding than another variety (e.g., Birmingham English). It has long been known that holding attitudes towards language can have serious discriminatory consequences (e.g., Purnell et al., 1999; Rickford & King, 2016; Fasoli & Maass, 2020; Lev-Ari & Keysar, 2010). However, one of the big unsolved questions about language attitudes is of cognitive nature: How do these attitudes emerge in the first place? It is possible to group potential factors triggering language attitudes into two areas of research that both deal with how language is being perceived and evaluated: *indexicality research* and *iconicity research*.

Indexicality research comprises approaches from social psychology and sociolinguistics (e.g., Lambert et al., 1960; Preston, 2017). Common to these approaches is that they assume language attitudes to be an *indexical* phenomenon. To the listener, linguistic features semiotically index (Peirce, 1958; Silverstein, 2003), or, in other words, point to groups of speakers. Listeners try to infer speakers' perceived geographical regions and social groups, their race and ethnicity, gender, sexual orientation, socioeconomic status, and character traits based on the

linguistic features they use (see, e.g., Lambert et al., 1960; Stewart et al., 1985; Munson & Babel, 2007; Preston, 2017). Crucially, in this indexical view, languages, varieties, and their linguistic features act as signifiers of attributes which are not related to the features themselves. Rather, linguistic features (e.g., German language features) merely point to groups of speakers (e.g., Germans). Only the association with specific groups of speakers then activates the corresponding evaluative meaning; for instance, a positive or negative aesthetic judgment ("German sounds harsh"). This view is encapsulated in what is called the *imposed norm hypothesis* (Giles et al., 1979), i.e., the idea that language attitudes result primarily not from language itself, but from differences in power and prestige among groups of speakers, and the *social connotations hypothesis* (Trudgill & Giles, 1976), i.e., the idea that language attitudes result from sociocultural stereotypes, cultural pressure, and social conditioning (also see Giles & Niedzielski, 1998).

Iconicity research (e.g., Kawahara et al., 2021; Winter et al., 2022) assumes at least some evaluative perceptions of language to be an *iconic* phenomenon. That is, to the listener, phonological markers possess qualities that cognitively connect them to the attributes with which they are associated. For example, the well-documented maluma-takete (Köhler, 1929) and bouba-kiki (Ramachandran & Hubbard, 2001) effects show that certain pseudowords are associated with round versus spiky shapes. Other prominent examples include the findings that /p/ and /b/ indicate fullness, /i/ indicates smallness (Blasi et al., 2016), that sibilants are associated with the action of flying, voiced obstruents are perceived as negative or "dark", and /p/ as "cute" (Kawahara et al., 2021), that swear words tend to lack approximants (Lev-Ari & McKay, 2022), and that there is a historically robust pattern of trilled /r/ being associated with rough textures (Winter et al., 2022). The list of effect types includes diverse dimensions such as size, shape,

action, brightness, speed, color, taste, and even (bordering on the territory of indexicality research) social dominance (see Barker & Bozic, 2024 for an overview). Crucially, in this iconic view, languages, varieties, and their linguistic features act as signifiers that are non-arbitrarily related to the attributes they are associated with. Evaluative semantics, for instance aesthetic judgments like "German sounds harsh", can be activated by how German sounds. For example, a language or variety that includes many voiced obstruents may be perceived negatively because it is articulatorily challenging to produce them (Kawahara et al., 2021). Or, the reason trilled /r/ is associated with roughness could be that its discontinuous phonetics resembles rough textures (Winter et al., 2022). This view echoes the *inherent value hypothesis*, which suggests that properties of linguistic features themselves play a role in aesthetic judgments. Figure 1 visualizes the difference between the indexical view and the iconic view on language attitudes. Whereas the indexical route assumes an indirect link between sound and meaning, with speakers in between, the iconic route assumes a direct link, for example, by way of shared resemblance.

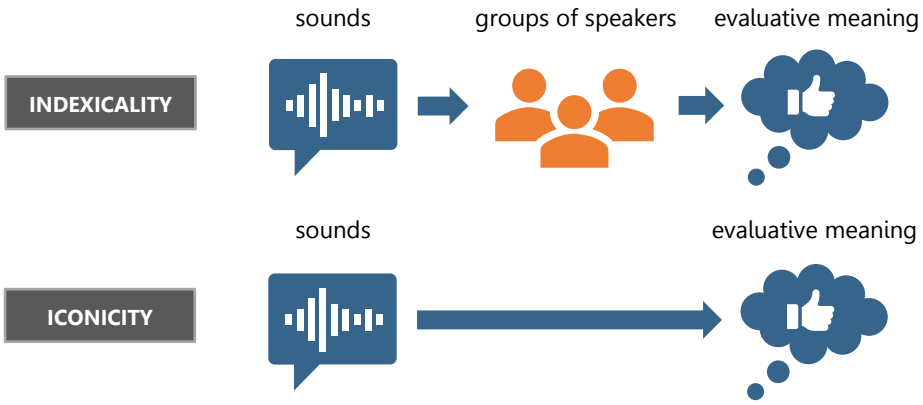


Figure 1: Simplifying overview of the indexical and the iconic link between sound and meaning.

Both hypotheses may contain truth, but how do we pit the two routes against each other empirically? One approach is to have participants rate varieties of languages unknown to them. Since the early days of language attitude research, some studies have reported that participants' ratings of these unknown varieties correspond to their social meaning (e.g., Ladegaard, 1998; Hilton et al., 2022), while others have reported the opposite (e.g., Giles et al., 1974; Giles et al., 1979). More recently, the renewed interest in the question has spawned studies that systematically analyzed ratings of many different languages according to sets of both phonetic-phonological and social variables: Reiterer et al. (2020) show that Central Europeans rate European languages differently on several semantic dimensions depending on social variables like familiarity with the language and L2 experience, and phonetic variables like sonority, vocalic share, and syllable structure. Anikin et al. (2023) find that a large number of languages are rated as more or less beautiful, across speakers of different languages, but not depending on sounds. Instead, some of the most important factors were perceived familiarity, voice quality, and different cultural biases. Finally, Mooshammer et al. (2023), instead of natural languages, used constructed languages (conlangs, languages created through conscious human engineering and invention, e.g., for use in fiction). They find that Germans rate several existing conlangs more positively (e.g., Sindarin and Quenya) or negatively (e.g., Klingon and Dothraki) depending on both their phonetic-phonological properties (e.g., voicing) and phonological familiarity.

There are three open issues that should be addressed following these studies. First, across studies, there is contradictory evidence, with some studies finding support for only phonetic-phonological predictors, some only for social predictors, and some for both. These results are also hard to compare, as they examined different semantic dimensions. For instance, Anikin et al. (2023) only looked at the beauty scale (ugly—beautiful). Second, these studies had listeners

rate entire languages, which vary in thousands of properties, and they do not attempt to isolate specific sounds or features. And third, the studies reviewed above tested existing languages (including existing conlangs), which makes it difficult to control for previous exposure. Anikin et al. (2023, p. 6) suggest that as an alternative to artistic conlangs like Elvish, it is also possible to use "meaningless synthetic speech" (also see a similar call by Hilton et al., 2022, p. 45). This would enable us to test whether we find similar patterns for language that is dissociated from social context.

The present study addresses these points by testing a variety of semantic differential scales with languages created from scratch that have never been heard by any listener. As a result, the experiment probes the iconic route by removing as much real-life noise as possible. In addition, this approach can isolate specific, potentially iconic phonetic-phonological features. To launch this endeavor, the study takes a close look at one of the most notorious candidates for supposedly making language sound unpleasant—the voiceless velar fricative /x/.

Why /x/? It turns out that this sound is one of the sounds that seem to have, plainly speaking, a bad reputation. /x/, and more generally a class of sounds that by laypeople is sometimes referred to as "gutturals," are often claimed to be perceived as negative, for instance, as sounding harsh. Mooshammer et al. (2023, p. 6), discussing Stanley (2003), highlight that so-called gutturals "are associated with growling, choking, hostility, and even illness." The class of gutturals is phonologically not always clearly defined but appears to be popular as a non-technical term among non-linguists. Generally, this latter, non-phonologically grounded use of the term encompasses sounds produced in or towards the back of the oral cavity, including velar, uvular, pharyngeal, or laryngeal consonants. The popular conception of guttural among English speakers, though, seems to exclude sounds English speakers are familiar with, like /k, g, ŋ,

h, ʔ/, but would include /x, ɣ, ɯ̟, ʟ, q, ɢ, ɴ, ʀ, χ, ʁ, ħ, ʕ, ɦ/. This makes the denotation of the term in non-academic registers somewhat different from its phonological definitions. For /x/ in particular, it is easy to find associations with harshness in online discussions, popular media, and memes, presumably because it prominently features in languages like German (the "*ach*-Laut") and Arabic, which are stereotypically framed as harsh. Where might these perceptions arise from?

Associations of growling, illness, and choking are intuitive due to the sound's shared resemblance to threatening animal sounds or coughing, respectively. The idea of hostility may be connected to the fact that velars or fricatives, i.e., sounds that are perceived as harsh or hissing, can indicate increased arousal and hostility among animals (Rendall & Owren, 2010; and discussed in the phonaesthetic context in Reiterer et al., 2020). In addition, there is evidence showing that compared to consonants articulated in the front of the mouth, those articulated in the back of the mouth are generally dispreferred among human listeners (e.g., Maschmann et al., 2020). As discussed by Hilton et al. (2022), this may be because, crosslinguistically, front consonants are ontogenetically acquired earlier than back consonants (Menn & Vihman, 2011). In iconicity research, general mappings of affective scales onto articulatory space are not unheard of. For instance, multiple studies have shown that the entire vowel space is categorized into positive and negative, or more pleasurable and less pleasurable (e.g., Körner & Rummer, 2022; also see the discussion in Winter et al., 2019). In short, /x/ is expected to contribute to more negative ratings of language.

1.2 Hypotheses

Taken together, we can expect the following. From the iconic perspective, form and meaning directly relate to each other, and iconic meaning is often taken to be "universal" or "inherent," for instance due to evolutionary hardwiring or physical resemblance in the sign. We can thus expect ratings to differ based on whether /x/ is present or not, regardless of sociocultural factors (such as language background, recognition of the language, or demographic information). Specifically, /x/ supposedly evokes negative associations. Hence,

$H1_{ICONICITY}$ The stimulus recordings with /x/ should be rated worse than the control recordings without /x/, regardless of sociocultural variables.

From the indexical perspective, however, associations are dependent on sociocultural factors. Meaning is unrelated to form itself, but created through social relations and groups. Such sociocultural factors capture our relations to and attitudes towards groups of people; they are shaped both by who these people are and by who we are—indexical meaning is constructed through social experience. We can thus expect ratings to differ based on sociocultural factors (such as language background, recognition of the language, or demographic information), regardless of the form itself (whether /x/ is present or not).

$H1_{INDEXICALITY}$ Ratings of the recordings should be affected by sociocultural variables (rather than by the speech signal itself).

Some of the variables that are commonly included are previous exposure, familiarity with a language, age, gender, and whether or how a language was recognized by the listeners (see, e.g., Reiterer et al., 2020, Anikin et al., 2023, Mooshammer et al., 2023, Hilton et al., 2022). How these and other predictors were implemented, and which specific directional hypotheses were

164 associated with them, is described in Section 2.5. But for all of these, while in the iconic view
165 what matters is not them but mostly the signal itself, in the indexical view what matters is not
166 the signal but these sociocultural predictors—how the listeners were brought up and socialized,
167 what they have experienced with their and other social groups, and what they associate with
168 specific groups of people, their region, identity, and supposed character traits. This direct con-
169 trast of iconic and indexical effects also suggests that we should examine how these variables
170 perform compared to each other. That is, in random forests (see Section 2.6), we expect that

$H2_{ICONICITY}$	Effects of the speech signal itself should be able to compete in predictive strength against sociocultural variables.
$H2_{INDEXICALITY}$	Sociocultural variables should outperform effects of the speech signal itself.

171 The following section explains in detail how these hypotheses were tested, including stimuli,
172 participants, questionnaire, but also the directional hypotheses and operationalizations for
173 each predictor and their statistical implementation.

174 2. Method

175 The experimental materials, data, and code are available in the OSF Repository.

2.1 Stimuli

For the creation of stimulus texts I wrote a Python program, the sonority-sensitive pseudotext generator, or SSPG. Unlike other generators (e.g., Rosenfelder, 2012), it automatically creates words with syllables that adhere to the sonority sequencing principle (in the present study, this is relevant only for adhering to the CV structure, see below). The SSPG can thus generate phonologically realistic samples of non-existing languages that control for many features that could affect attitude ratings (see Table 1), allowing the researcher to only vary the feature of interest, in this case /x/.

I generated three texts for the control condition and three texts for the stimulus condition, to be spoken by different speakers. Both conditions share the same baseline phonemic inventory /p, t, k, b, m, n, s, l, w, j/. The only difference between the control condition and the stimulus condition is that in addition, only the stimulus condition also contains /x/. In this condition, the sound /x/ occurs on average in about 20 out of the 100 words in the text, which was judged to be a good balance between sufficient salience and realism. The remaining properties were held constant (see Table 1 for an overview).

Table 1: Overview of properties of texts in both conditions, stimulus and control, after verification. The sound inventory *base* refers to the baseline inventory /p, t, k, b, m, n, s, l, w, j, i, u, a, e, o/. The crucial difference between control and stimulus conditions is the /x/ added to the baseline inventory.

Condition	Control			Stimulus		
Text ID	1	2	3	1	2	3
Mean sonority 1–17	10.91	10.94	10.92	10.94	10.95	10.92
Consonants %	47.62	46.9	47.26	47.33	47.44	47.13
Obstruents %	25.32	23.97	23.73	24.52	26.07	26.43
Vowels %	52.4	53.1	52.7	52.7	52.6	52.9
Voicing %	79.44	80.37	79.51	79.53	78.21	78.28
Syllable structures	CV, V	CV, V	CV, V	CV, V	CV, V	CV, V
Syllable weights 0–1	0.9, 0.1	0.9, 0.1	0.9, 0.1	0.9, 0.1	0.9, 0.1	0.9, 0.1
Number of words in the text	100	100	100	100	100	100
Maximum number of syllables	4	4	4	4	4	4
Sound inventory	base	base	base	base, /x/	base, /x/	base, /x/

194

195 To generate audio files from these texts, I used the speech synthesis software Amazon
196 Polly. I added SSML tags to the texts (Speech Synthesis Markup Language, Baggia et al., 2010)
197 telling the speech synthesizer to pronounce them exactly as specified in the IPA transcription
198 provided, rather than interpret them as orthographic transcriptions. Further SSML tags were
199 added specifying when to adapt intonation to sentence boundaries. One challenge is that speech
200 synthesis software is designed to work language-specifically. In other words, it is optimized for
201 a certain language. There is no "neutral" way to realize a string of phonemes, neither for hu-
202 mans nor for software. This has two consequences. First, a given voice in Polly, or any other
203 available speech synthesis software, will not be able to realize all existing human sounds. This
204 makes it difficult to synthesize newly constructed languages. The solution is to use phonemic
205 inventories with sounds that are crosslinguistically very frequent, as introduced above, and thus
206 more likely to be covered by many voices. Second, even if the phonemic input is the same across
207 voices, each voice will be optimized for specific phonetic realizations and suprasegmental fea-
208 tures, such as intonation. One option to address part of this problem is to remove supraseg-
209 mental features altogether (e.g., to flatten the pitch in the stimulus). This, however, does not
210 address small differences in the realization of phonemes and will make the stimulus sound ex-
211 tremely artificial. The alternative, then, is to embrace the reality of different language optimi-
212 zations consciously by adopting several ones and holding each one constant between condi-
213 tions. This study followed the latter strategy.

214 To cover a varied spectrum of optimizations that can be statistically controlled for, this
215 study used voices from five different languages, which are Arabic, Dutch, German, Polish, and
216 Spanish. These five were chosen for two reasons. The first reason, mentioned above, was that

the voices needed to be able to convincingly realize the stimuli, including /x/, meaning both the baseline inventory and /x/ had to overlap with the inventory of the language for which the voice was optimized. The second reason was the availability of different voices in the language. I only considered languages that had at least three voices per language, and that had at least one female voice and one male voice. They also had to cover different kinds of technology across voices. Each language should be represented by at least one neural engine voice, which produces more realistic results, and one standard engine voice, which features less smart transitions between sounds. For every language, the six texts were pseudo-randomly distributed over three voices, so that every voice read one text from each condition. This amounted to 30 recordings in total. Table 2 provides an overview of all voices.

Table 2: Overview of Amazon Polly voices used to create the recordings, together with their language optimizations. C = control condition, S = stimulus condition.

Arabic						Dutch						German						Polish						Spanish					
Hala		Zayd		Zeina		Laura		Lotte		Ruben		Daniel		Marlene		Vicky		Ewa		Jacek		Ola		Conchita		Lucia		Sergio	
C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S

2.2 Participants

501 participants were sampled on Prolific. A quota sampling by PRIMARY LANGUAGE was used to test for effects of previous exposure to /x/. This will be explained in detail in Section 2.5 (predictor EXPOSURE). 277 listeners were male, 213 female, and 11 non-binary. Age ranged from 18–72 (mean = 32, median = 29). Participants gave their informed consent and were financially compensated.

2.3. Questionnaire

The questionnaire was constructed using SoSci Survey (Leiner, 2024). After the information and consent sheet, listeners were presented with a story about robots. This was done for two reasons, first, to ensure participants would find the artificially synthesized voices believable, and second, to make sure participants did not carry over any previous non-linguistic associations with specific groups of people. The story, however, emphasized that the robots are human-like. This ensured at least a basic level of social relevance, which could be considered important for indexical effects to emerge (see discussion in Section 4).

The next page informed participants that they would now listen to the first of three pairs of robots. A pair always consisted of a stimulus audio file and a control audio file (in randomized order, with language optimization as between- and voice as within-participants factor). Each recording was followed by the rating scales (see Section 2.4), which were implemented as sliders from 0 to 100. After the scales, listeners had to rate how familiar this language sounded to them, and type into a free text field which real language or dialect they thought this language resembled the most. Then, the next stimulus of the robot pair was presented. Each participant listened to three pairs, i.e., six recordings.

2.4 Response variables

Ten semantic differential scales were chosen for the present study. These scales are taken to represent different underlying dimensions, based on the literature (e.g., Giles & Niedzielski, 1998; Reiterer et al., 2020; Hilton et al., 2022). The measurements on these scales serve as response variables in the statistical models. Table 3 provides an overview.

Table 3: Overview of semantic differential scales used as response variables in the experiment, together with their negative valence labels (numerical value 1) and positive valence labels (numerical value 100). The final column indicates which scales are used to represent which underlying dimension identified from the literature.

Scale	negative valence 1		positive valence 100	dimension
PLEASANTNESS	unpleasant	—	pleasant	general affinity
BEAUTY	ugly	—	beautiful	aesthetics
SOFTNESS	hard	—	soft	aesthetics
SHAPE	spiky	—	round	aesthetics
EDUCATION	uneducated	—	educated	status
INTELLIGENCE	stupid	—	intelligent	status
FRIENDLINESS	unfriendly	—	friendly	solidarity
ORDINARINESS	strange	—	normal	solidarity
GOODNESS	evil	—	good	morality
EROTICISM	unerotic	—	erotic	eros

The most general underlying dimension is what we could call *general affinity*, which is usually represented by generic scales like likeability (do not like at all—like very much, Anikin et al., 2023) or, as in this case, PLEASANTNESS (unpleasant—pleasant, Giles & Niedzielski, 1998; Reiterer et al., 2020).

More specific than general affinity, it is possible to identify in the literature a dimension we could call *aesthetics*, which can be used to group scales like elegance (inelegant—elegant), fashion (unfashionable—fashionable), sweetness (harsh—sweet), melody (tuneless—melodic), or orderliness (chaotic—orderly) (see, e.g., Giles & Niedzielski, 1998, Reiterer et al., 2020, Hilton et al., 2022). For the present study, this dimension is represented by BEAUTY (ugly—beautiful) as it is perhaps the most general and widely used scale in this category (Anikin et al., 2023; Giles & Niedzielski, 1998; Hilton et al., 2022; Reiterer et al., 2020, but see Domizi, 2024 and Lopes, 2017 on the difference between likeability and aesthetics). In addition, I selected SOFTNESS (hard—soft) and SHAPE (spiky—round) as the two perhaps most prominent semantic differentials in iconicity research (cf. the bouba-kiki effect, Ramachandran & Hubbard, 2001).

The two most prominent dimensions in indexicality research are *status* and *solidarity*. They can be represented by a variety of scales (see, e.g., Stewart et al., 1985, Giles and Niedzielski 1998, Bayard et al., 2001, Clopper & Pisoni, 2004, Reiterer et al., 2020, among many others), such as, for status, importance (marginal—influential), culture (uneducated—cultured), wealth (poor—wealthy), success (unsuccessful—successful), or competence (incompetent—competent); for solidarity, kindness (unkind—kind), fun (boring—fun), coolness (uncool—cool), politeness (impolite—polite), generosity (stingy—generous), welcomingness (repelling—welcoming), or comfortableness (uncomfortable—comfortable). Here, status is represented by the classics EDUCATION (uneducated—educated) and INTELLIGENCE (stupid—intelligent), while solidarity is represented by FRIENDLINESS (unfriendly—friendly) and ORDINARINESS (strange—normal, cf. Hilton et al., 2022).

Studies on conlangs in particular (Mooshammer et al., 2023) have highlighted the relevance of a *moral* dimension, since some conlangs are constructed with the design goal in mind of conveying evilness to the audience. This dimension can be represented by peacefulness (aggressive—peaceful) and solidarity scales like trustworthiness (untrustworthy—trustworthy), honesty (dishonest—honest), or reliability (unreliable—reliable), and here is represented by the most straightforward one, GOODNESS (evil—good).

Finally, the studies by Reiterer et al. (2020) and Kogan and Reiterer (2021) have emphasized that eros, too, is a dimension in which listeners readily rate language. This is connected to the so-called Latin Lover effect, by which Romance languages are perceived as sensual, attractive, or sexy-sounding. Viable proxies for eros are romanticism (unromantic—romantic), seductiveness (unseductive—seductive), sexiness (unsexy—sexy), or attractiveness (unattractive—attractive). Here, I use the presumably most basic scale EROTICISM (unerotic—erotic).

297 Overall, the scales selected for this study constitute a set considered to tap into the most relevant
298 and different underlying semantic dimensions. Each participant provided six ratings per scale,
299 which amounts to 3006 ratings per scale overall.
300
301 RATING
302 In addition to the individual scales introduced above, one response variable in this study is
303 RATING (negative—positive). This variable combines the ratings of each scale according to their
304 negative-positive alignment in a long-form dataset. This is useful to investigate a general ten-
305 dency in whether the stimuli with /x/ are rated worse (or better) than the control.
306 It is worth noting that doing this raises questions about the ontology of the semantic
307 scales. While for some scales, their alignment with the negative-positive axis is obvious (for
308 instance, unpleasant is negative, pleasant is positive), for others, it may not be as intuitive (for
309 instance, it may not be self-evident that spiky is more negative than round, or that soft is more
310 positive than hard). A few studies demonstrate the relative independence of scales. For example,
311 O'Connor and Barclay (2017) find that effects of pitch on perceptions of trustworthiness are
312 independent of pitch effects on attractiveness. And, even though high status and high solidarity
313 could both be considered as having positive valence, a classic finding in much of the sociolin-
314 guistic literature is that the solidarity dimension tends to negatively correlate with the status
315 dimension (see Kircher & Zipp, 2022 for an overview).
316 However, there is also evidence that even scales that would not necessarily be expected to
317 be underlyingly positive or negative are indeed interpreted this way. For example, for shape,
318 round is positive and spiky is negative, either because spiky shapes are threatening (Bar & Neta,
319 2006) or because round shapes by themselves are inherently likeable (Vartanian et al., 2013;

Bertamini et al., 2015; Palumbo et al., 2015; for a discussion, see Domizi, 2024, pp. 82–85). Similarly, soft and smooth is better than hard and rough (e.g., Essick et al., 2010; Essick et al., 1999). As Winter et al. (2019, p. 4) summarize, pleasantness is often a "common denominator," and there is more and more "evidence for the presence of affective iconicity in language, i.e., the association between speech sounds and pleasant or unpleasant feelings." For the present study, I performed a reality check, making sure that ratings on all scales correlated in the expected direction (i.e., according to their hypothesized negative-positive valence). All these correlations played out as expected, indicating that the negative-positive alignment as shown in Table 3 is supported by the data.

2.5 Predictor variables

CONDITION

This is the main predictor of interest, coding whether the audio file was a `stimulus` recording including /x/ or a `control` recording without /x/.

EXPOSURE

This predictor coded for how exposed listeners had been to /x/ before participating in the experiment. For `exposed` listeners, I sampled participants who had indicated as their primary spoken language one of the five languages with /x/ from the voice's language optimization, i.e., Arabic, Dutch, German, Polish, or Spanish (each 10 % or 50 listeners). For `less exposed` listeners, I sampled participants who had indicated as their primary spoken language English, Italian, or Japanese (each about 17 % or 86/85/80 listeners respectively). An overview is given

in Table 4. Note that the predictor codes for whether or not /x/ occurs in the language, disregarding phonotactic position or allophony (see discussion in Section 4). We can expect that any potential iconically motivated aversion to /x/ will be cushioned by the listener being used to that sound (also see the rationale for FAMILIARITY below).

Table 4: Overview of participant numbers according to their exposure to /x/ and primary language.

Exposure	exposed					less exposed		
Primary language	Arabic	Dutch	German	Polish	Spanish	English	Italian	Japanese
Number of participants	50	50	50	50	50	86	85	80
Total	250					251		

RECOGNITION

RECOGNITION specifies which real group of languages listeners thought the recording was most similar to. I tested three ways of grouping the listeners' answers: (1) as a binary variable simply encoding whether the listener provided any guess or not, (2) as areas, such as East Asia, South East Asia, Eastern Europe, etc., and (3) as a mix of larger macroareas and groups for which there are specific hypotheses. I considered the first option to be too uninformative, and the second one to have too many values, which can lead to statistical problems. The most sensible solution was the third option.

The possible values recognition could take were `Africa`, `Asia`, `Germanic`, `Middle East`, `Romance`, and `other`. While Africa and Asia are both macroareas, the Middle East refers to a more specific region, while Germanic and Romance refer to specific language families. As mentioned above, the reason for coding these more specific values is that we have concrete hypotheses associated with them. As described in Section 1, languages from the Middle East, like Arabic, and Germanic languages, like German, seem to be notoriously framed as

"harsh," "unpleasant," or "aggressive." Recordings with which listeners associated a label coded as `Germanic` (like "German" or "Germanic") are thus expected to be rated comparatively more negative overall, particularly in terms of pleasantness and aesthetics (except perhaps for status, for a discussion of the complex situation for the example of the German language see Domizi, 2024). Note that `Germanic` was deliberately coded such that it does not include English, as English is usually not stereotypically described with these negative attributes. Romance languages, on the other hand, are notoriously perceived as "beautiful," "romantic," or even "erotic," at least in Eurocentric discourse (see again studies on the Latin Lover effect, like Reiterer et al., 2020). Recordings with which listeners associated a label categorized as `Romance` (like "French" or "Italian") were thus expected to be rated comparatively more positive, especially on scales related to aesthetics and attractiveness. `Other` included cases where no label was provided, unclear and ambiguous labels (e.g., "Eastern," "mixed"), and cases that did not fit the categories of interest (e.g., "Native American," "Pacific Islands").

FAMILIARITY

We like that which we know. There is no shortage of either evidence or explanations for this phenomenon—perhaps familiarity reduces fear of harm (Bornstein, 1989) or facilitates processing (Reber et al., 2004)—and it should and does apply to language (for a discussion, see Kogan & Reiterer, 2021). Familiarity with the language was elicited as self-reported familiarity judgements by the question "How familiar does this language sound to you?" followed by a slider on a scale from 1 (not at all familiar) to 100 (very familiar).

LANGUAGE

This variable specifies the language for which the respective Amazon Polly voice was optimized, and can thus take the values `Arabic`, `Dutch`, `German`, `Polish`, or `Spanish`. This captures differences in suprasegmental features and voice quality. As not much is known about attitudes towards specific patterns of intonation, stress, or rhythm, we should refrain from formulating specific hypotheses for this predictor.

GENDER OF VOICE

The gender of the Amazon Polly voice used to synthesize the texts to speech. It can take the values `male` or `female`. In general, humans prefer voices that sound sex-typical (Pisanski & Feinberg, 2018). However, it is also possible to expect differences in the rating of these two genders. One hypothesis could be that in the current patriarchal system, female voices score better on solidarity scales (due to metaphors of women as social, nurturing, caring, etc.), while male voices could score better on status (due to associations with authority, success, power, etc.). This effect may be augmented due to lower pitch being associated with dominance (see, e.g., Calhoun et al., 2024; Borkowska & Pawlowski, 2011). Female voices could also be hypothesized to be rated better on the eroticism scale, as women are more pervasively sexualized, and on the goodness scale, as male voices are on average lower-pitched and may thus activate the "evil sounds deep" trope (for a collection of examples, see tvtropes.org/pmwiki/pmwiki.php/Main/EvilSoundsDeep). Overall, we should find a preference for female voices, as these were rated better in previous studies on the aesthetics of languages (Anikin et al., 2023; Reiterer et al., 2020) and have been shown by neurophysiological studies to produce stronger activation patterns in the brain (e.g., Lattner et al., 2004). Better ratings of female voices may also be related to pitch, as higher frequency is perceived as less

dark and thus presumably better than lower frequency (Marks, 1975; but see Mooshammer et al., 2023).

GENDER OF LISTENER

The gender of the listener, which in this dataset takes the values `female`, `male`, and `non-binary`. We can expect GENDER OF LISTENER to interact with GENDER OF VOICE, for instance because listeners (especially women, Rudman & Goodwin, 2004) may rate voices better that correspond to their own gender (in-group bias, see Giles & Johnson, 1987; Hewstone et al., 2002). Assuming that most of the participants are heterosexual, one could expect the opposite effect for the scale eroticism. Here, male listeners may rate female voices as more erotic, while female listeners may rate male voices as more erotic. In addition, we can expect to replicate a general effect between men and women where women rate language varieties that are different from their own better than men (Coupland & Bishop, 2007).

OTHER VARIABLES

Other covariates included in the models were POLYGLOT (the listener's cumulative language proficiency, measured as the sum of the number of L1s and L2s, Reiterer et al., 2020), AGE, MUSICALITY (the musical experience of the listener on a scale from 1 to 10), LINGUISTICS (previous experience with linguistics, `yes` or `no`), INPUT (input device by which participants accessed the experiment), OUTPUT (output device participants used to listen to the recordings), LOCATION (location in which participants carried out the online experiment), SCALE, (the semantic scale), and PARTICIPANT (as a random intercept).

2.6 Modeling

The data were modeled with linear mixed-effects models, using the `lme4` package (Bates et al., 2015) and `lmerTest` (Kuznetsova et al., 2016) in `R` (R Core Team, 2025). One model was created to model ratings from all scales simultaneously in one concatenated dataset, with the purpose of investigating a general negative-positive tendency in the evaluation of the stimuli. This model had RATING as the response variable and will be referred to as the *across-scale model*. It was structured as follows:

(1) RATING ~ CONDITION * EXPOSURE + RECOGNITION + FAMILIARITY + LANGUAGE + GENDER OF LISTENER * GENDER OF VOICE + POLYGLOT + AGE + MUSICALITY + LINGUISTICS + INPUT + OUTPUT + LOCATION + SCALE + (1 | PARTICIPANT)

As can be seen from this model structure, two pairs of predictors were included as interactions. First, EXPOSURE needed to be allowed to interact with CONDITION, since we would assume that only the stimulus would be affected by previous exposure to /x/. Second, GENDER OF VOICE should interact with GENDER OF LISTENER, as ratings could be affected by both an in-group bias and by sexuality (see Section 2.5).

In addition to the across-scale model, I also created one model for each semantic differential scale. For these models, the response variable was the rating on one specific scale, for instance, PLEASANTNESS. These models were identical to the across-scale model except for the fact that they did not include SCALE as a predictor.

Post-hoc comparisons of multiple contrasts of categorical predictors were computed with `emmeans` (Lenth & Piaskowski, 2026) using Tukey adjustment. In order to better compare the social variables against the phonological predictor, I also built random forests using the `randomForest` package (Liaw & Wiener, 2002), each with 500 trees. These featured the same

variables used for the linear mixed-effects regression models, without the interactions. The increase in node purity in the random forests can be used to estimate how important the predictors are compared to each other in predicting, or splitting, the data.

3. Results

First, comparing the distributions for each scale of the stimulus condition against the control condition (Figure 2), we see that the two conditions differ only slightly but overall have remarkably similar shape. This observation already tells us that whether /x/ is present or not does not seem to make a huge difference. This is impressive given that the task was designed in a way that put stimuli and control recordings in direct opposition, in an attempt to make it possible for listeners to compare them.

But let us now look at the output of the statistical modeling. Let us start with the across-scale model, before going into detail for relevant individual scales. Table 4 summarizes the results from this model. For reasons of space, the summary omits the control variables INPUT, OUTPUT, LOCATION, and SCALE. None of these showed significant effects except for SCALE (for instance, the eroticism scale yields significantly more negative ratings), which is expected but not central to the discussion. As for the remaining covariates, neither LINGUISTICS, nor MUSICALITY, AGE, or POLYGLOT show any effects [1]. LANGUAGE shows that the Arabic voice

[1] Following the suggestion of a reviewer, I also checked for interaction effects between musicality and condition. They suggest that more musical listeners might perceive /x/ as an acoustic texture, or timbre, rather than a social signal. The interaction itself is significant at $p < .05$ such that musicality has a slightly stronger positive effect on ratings in the stimulus condition compared to the control condition. However, none of the two conditions show an effect of musicality as such, i.e., at this point we are unable to find support for musicality affecting ratings of either the control samples or the stimulus samples themselves. This mirrors the lack of effects for musicality for both conditions when included as a main effect.

optimizations were rated significantly better than Dutch, and an analysis of all contrasts reveals that the Arabic optimization was indeed rated significantly better than any other optimizations (marginally significantly for German), while all other comparisons remain non-significant.

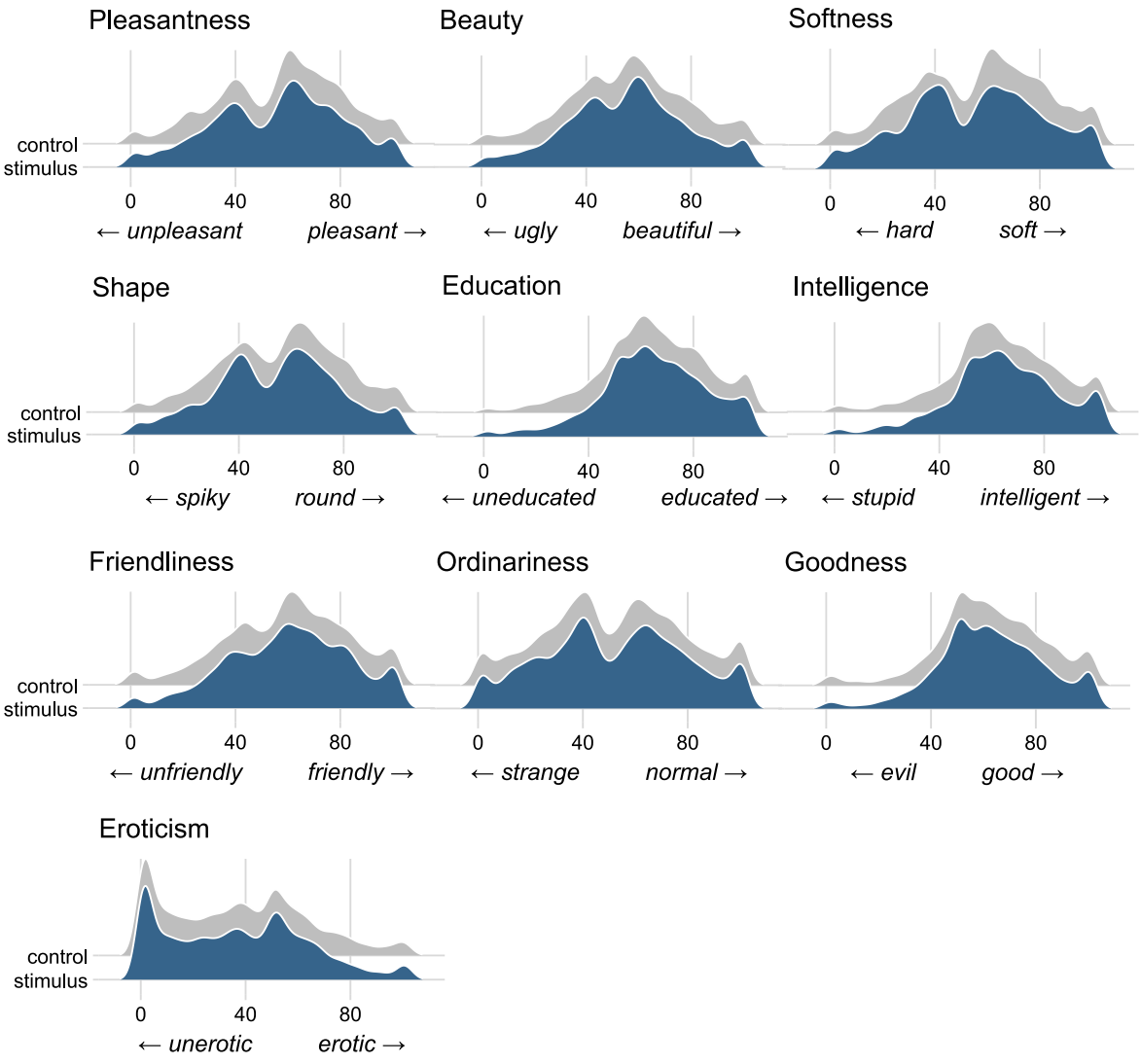


Figure 2: Distributions of ratings by condition, with one panel for each scale. The darker shaded distributions represent the stimulus condition with /x/, the lighter shaded distributions represent the control condition without /x/.

Table 5: Summary of across-scale linear mixed-effects regression model regressing the predictors described in Section 2.5 against RATING, with INPUT, OUTPUT, LOCATION, and SCALE omitted. Reference levels are EXPOSURE `less exposed`, CONDITION `control`, RECOGNITION `Africa`, LANGUAGE `Dutch`, GENDER OF LISTENER `female`, GENDER OF VOICE `female`, and LINGUISTICS `no`. The full model, as well as model outputs of regression models for individual scales, are documented in the OSF Repository.

Predictors		Estimate	Confidence interval			p-value	
Intercept		53.65	48.04	—	59.26	<0.001	***
EXPOSURE	`exposed`	−2.60	−4.81	—	−0.39	0.021	*
	`unclear`	−0.73	−4.44	—	2.97	0.698	
CONDITION	`stimulus`	0.25	−0.40	—	0.90	0.447	
RECOGNITION	`Asia`	0.44	−0.54	—	1.41	0.379	
	`Germanic`	−3.93	−5.41	—	−2.44	<0.001	***
	`Middle East`	−4.29	−5.61	—	−2.96	<0.001	***
	`Romance`	−3.41	−4.39	—	−2.43	<0.001	***
	`other`	−3.50	−4.48	—	−2.53	<0.001	***
FAMILIARITY		0.28	0.26	—	0.29	<0.001	***
LANGUAGE	`Arabic`	3.97	0.84	—	7.11	0.013	*
	`German`	1.02	−2.12	—	4.16	0.523	
	`Polish`	−0.92	−4.05	—	2.20	0.563	
	`Spanish`	−1.78	−4.88	—	1.31	0.259	
GENDER OF LISTENER	`male`	−4.32	−6.41	—	−2.22	<0.001	***
	`non-binary`	0.13	−6.82	—	7.08	0.970	
GENDER OF VOICE	`male`	−2.88	−3.60	—	−2.17	<0.001	***
POLYGLOT		−0.59	−1.22	—	0.03	0.061	.
AGE		0.01	−0.09	—	0.10	0.898	
MUSICALITY		0.22	−0.14	—	0.59	0.230	
LINGUISTICS	`yes`	0.74	−1.39	—	2.86	0.496	
$\text{EXPOSURE}_{\texttt{exposed}}$	× $\text{CONDITION}_{\texttt{stimulus}}$	−1.30	−2.23	—	−0.38	0.005	**
$\text{EXPOSURE}_{\texttt{unclear}}$	× $\text{CONDITION}_{\texttt{stimulus}}$	−0.48	−2.05	—	1.10	0.553	
$\text{GENDER}_{\texttt{male}}$	× $\text{VOICEGENDER}_{\texttt{male}}$	−0.62	−1.56	—	0.33	0.203	
$\text{GENDER}_{\texttt{non-binary}}$	× $\text{VOICEGENDER}_{\texttt{male}}$	−2.10	−5.32	—	1.12	0.201	

Random effects			
σ^2	374.83	ICC	0.24
$\tau_{00\ \text{participant}}$	116.48	$N_{\text{participant}}$	501
Observations	30060	Marginal / Conditional R^2	0.223 / 0.407

Let us now turn to our phonological predictor of interest, CONDITION, as well as to the social predictors EXPOSURE, FAMILIARITY, RECOGNITION, and GENDER.

We first examine CONDITION in interaction with EXPOSURE. From the interaction in Table 4 that shows the effect of the `stimulus` compared to the reference level `control` in the

`exposed` group, we can observe a significant negative effect on the ratings. To examine the full picture with all levels, Figure 3 plots this interaction effect together with the significance level for all contrasts (Tukey adjusted). We disregard `unclear` for the moment, as we did not have any predictions for this category.

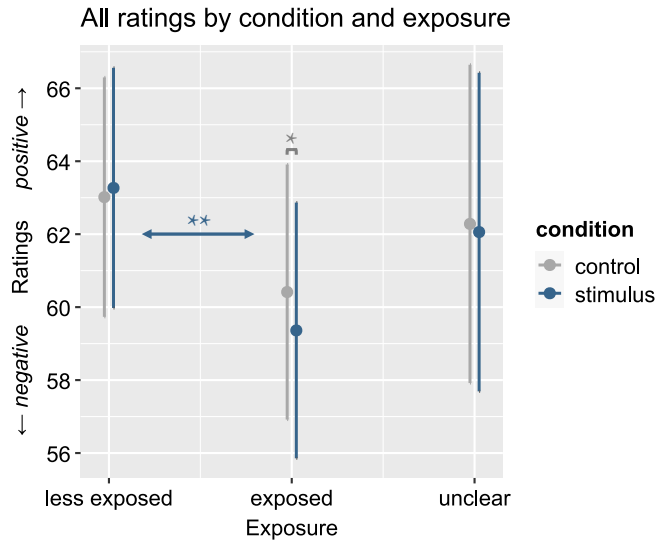


Figure 3: Interaction between CONDITION and EXPOSURE in the across-scale model. Significance codes: *** for $p < .001$, ** for $p < .01$, * for $p < .05$, no asterisks for $p > .05$, Tukey adjusted. Error bars indicate 95% CIs.

First, we find that exposed listeners rate the language with /x/ as significantly worse than the language without /x/, whereas we find no such difference in the less exposed group. If the negative effect of /x/ was purely iconic, it should have manifested in both groups of listeners regardless of exposure. From an iconic perspective, we could also have expected the exposed group to rate the stimulus better than the control, perhaps because this group is more used to /x/, which could mitigate its inherently negative associations. Interestingly, we find the opposite. If a listener had been exposed to /x/, it sounds worse to them. To foreshadow the discussion, one explanation comes from the indexical view: exposed listeners are likely more aware of the

respective stereotypes surrounding /x/, and thus rate languages with /x/ worse despite being more used to them. We can confirm this picture from a different angle if we compare exposure groups by condition. We find that exposed listeners rate the language with /x/ significantly worse than less exposed listeners.

The pattern of exposed listeners rating the stimulus worse than the control, and exposed listeners rating the stimulus worse than less exposed listeners, is indeed a pattern on all scales, though the effects do not always reach significance. This can be gleaned from the overview in Figure 4, which displays the same plot but for each statistical model, including the models for individual scales. To summarize, considering only the significant effects (see Table 5 for an overview of significances), we can say that compared to the control condition, exposed listeners rate the language with /x/ as less positive; and compared to less exposed listeners, exposed listeners rate the language with /x/ or its speakers as less positive, uglier, less erotic, less educated, and less intelligent (the final two dimensions also apply to their ratings of the language without /x/).

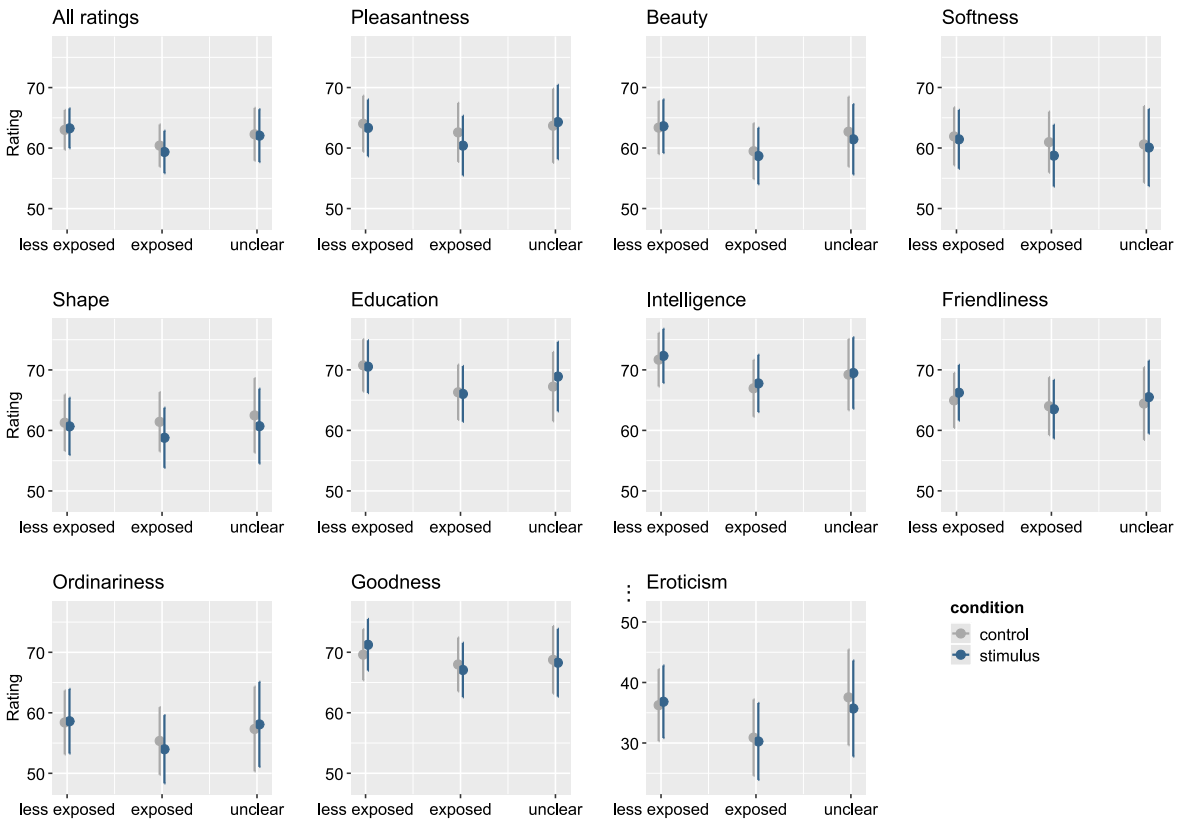


Figure 4: Overview of the interaction between CONDITION and EXPOSURE in all models. Rating on the y-axis is always arranged so that higher values are more positive. Error bars indicate 95% CIs. Note that the range displayed is the same for all panels except for EROTICISM, which received lower ratings overall.

Let us now move on to the remaining predictors. FAMILIARITY is of particular interest, as previous studies have used this measure to assess the indexical route, or social influence. FAMILIARITY turns out to be one of the strongest predictors in the across-scale model, as well as the most consistently strong predictor across all models. We see a very clear positive, and expected, effect across all scales (Figure 5). More self-reported familiarity always, and always very highly significantly, leads to better ratings, no matter the semantic differential scale.

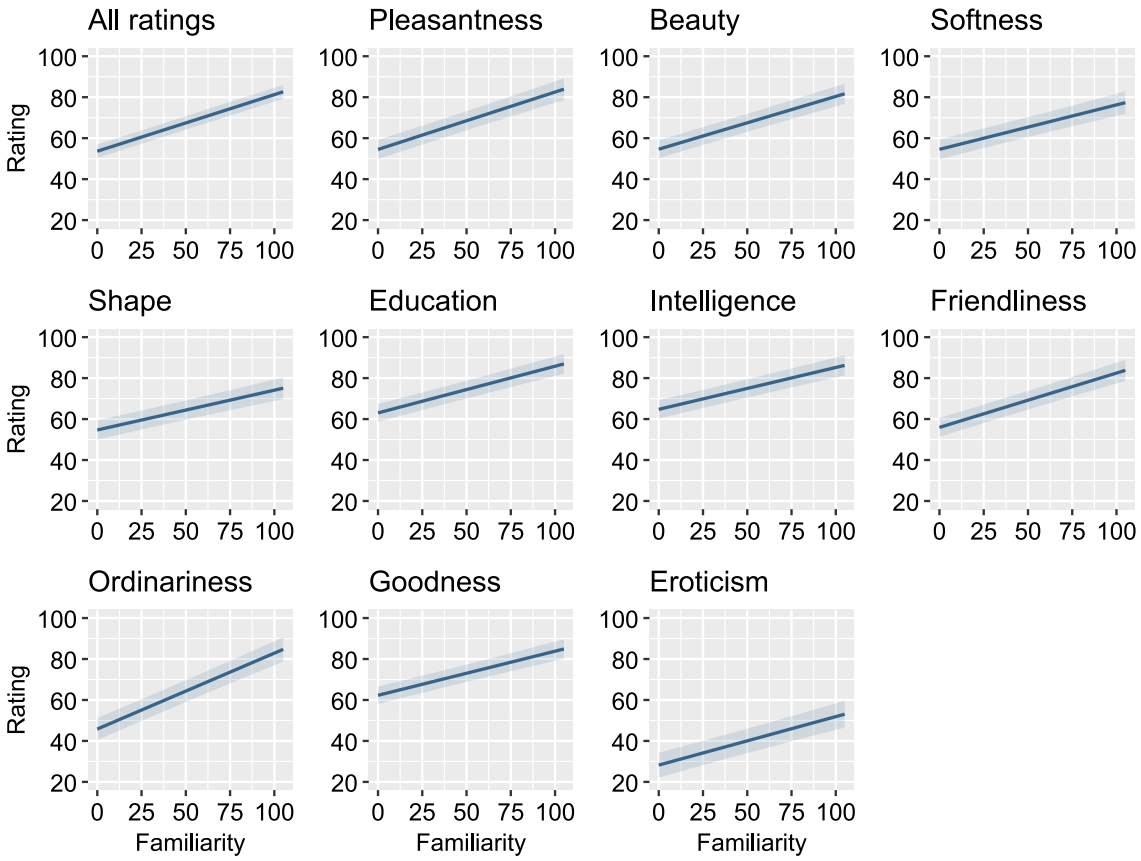
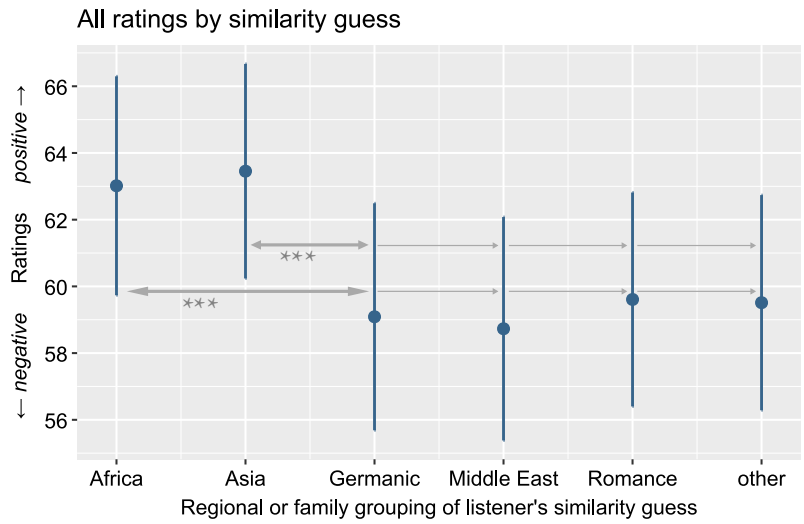


Figure 5: Overview of the effects of FAMILIARITY in all models. Rating on the y-axis is always arranged so that higher values are more positive. All effects are very highly significant at $p < .001$.

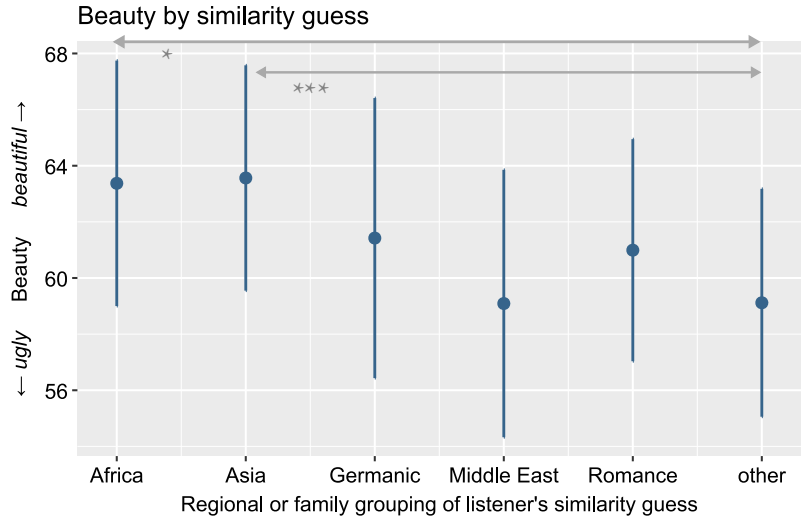
We now turn to RECOGNITION. As described in Sections 2.3 and 2.5, listeners were asked which language or dialect they thought the respective robot language resembled the most. In Table 4, we see that compared to the reference level `Africa`, listeners rate the languages worse if they think they are similar to a Germanic language (note again that Germanic does not include English here), a language from the Middle East, or a Romance language. The top panel of Figure 6 plots these effects. The negative ratings for Germanic languages and languages from the Middle East are expected. These are stereotypically framed as sounding "harsh," "unpleasant," even "aggressive," and these score the worst. However, Romance languages score only a little higher, which is surprising. Recall that the reason to create a category "Romance" was

primarily due to scales BEAUTY and EROTICISM, as I had expected to replicate the Latin Lover effect.

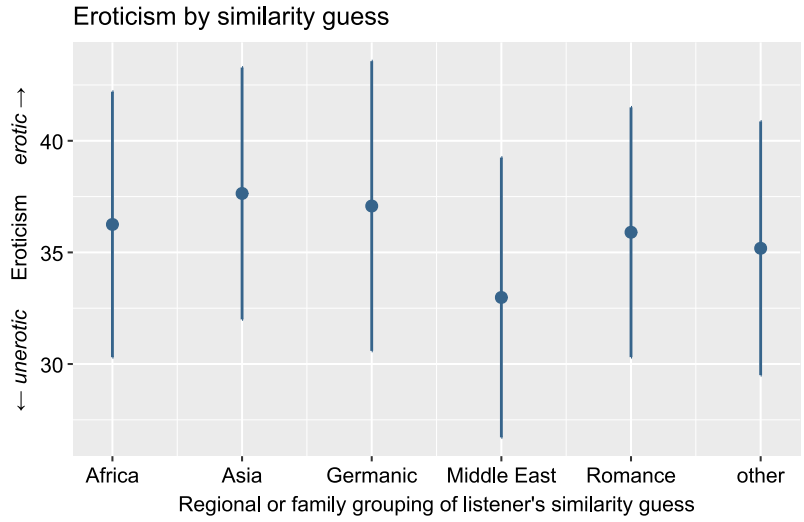
To investigate this, let us move away from the across-scale model and look at the BEAUTY and EROTICISM models in isolation. The middle panel in Figure 6 shows that for BEAUTY, only the category `other` scores significantly worse compared to Africa and Asia. Meanwhile, Romance languages are still perceived as somewhat beautiful, or at least non-significantly less beautiful than the other groups. Moving down to the bottom panel of Figure 6, a similar picture emerges for the crucial scale EROTICISM. When listeners think they hear a Romance language, and rate how erotic the language sounds, the negative effect for Romance languages from the across-scale model disappears. It is, however, worth noting that listeners also fail to rate perceived Romance languages as *more* erotic than the other groups. It is thus not possible to claim to have found evidence for the Latin Lover effect.



550



551



552

553 Figure 6: Overview of the effects of RECOGNITION for the across-scale model (top panel), the BEAUTY model (mid-
554 dle panel), and the EROTICISM model (bottom panel). Significance codes: *** for p < .001, ** for p < .01, * for p < .05,
555 no asterisks for p > .05, Tukey adjusted. Double arrows indicate a contrast; extending single arrows indicate that
556 significance also applies between the reference level and the following contrasts. Error bars indicate 95% CIs.

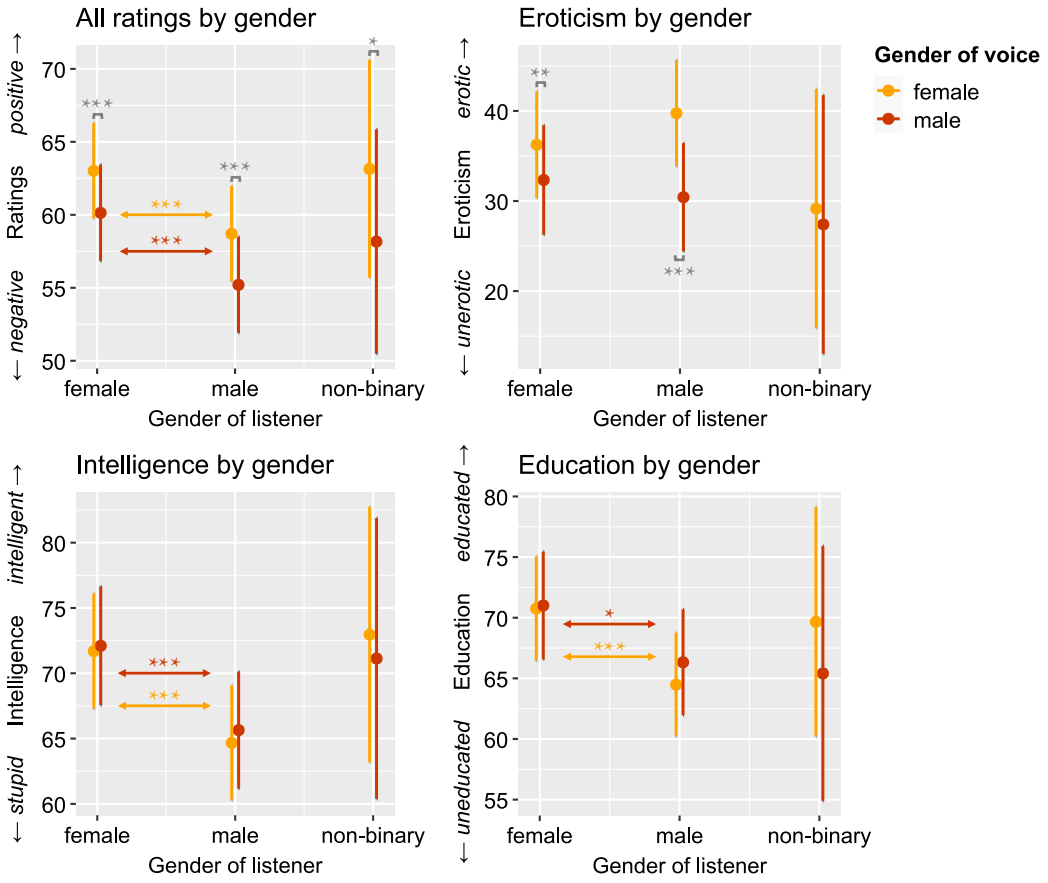


Figure 7: Overview of the effects of GENDER OF LISTENER in interaction with GENDER OF VOICE for the across-scale model (top left panel), the eroticism model (top right panel), the intelligence model (bottom left panel), and the education model (bottom right panel). Significance codes: *** for p < .001, ** for p < .01, * for p < .05, no asterisks for p > .05, Tukey adjusted for post-hoc comparison of multiple contrasts. Error bars indicate 95% CIs.

Let us now move to the final important social predictor, which is GENDER. From the model output in Table 4 we can observe two negative effects of both male voices and male listeners in the respective reference levels (`female` for each), but interaction effects are more easily understood as a plot (top left panel of Figure 7). Two things are of interest. First, we see that female, male, and non-binary listeners rate male voices as significantly worse than female voices. There is a general preference for female voices. Second, we see that male listeners rate

both female and male voices as more negative than female listeners. In other words, men generally rate everything as more negative than women.

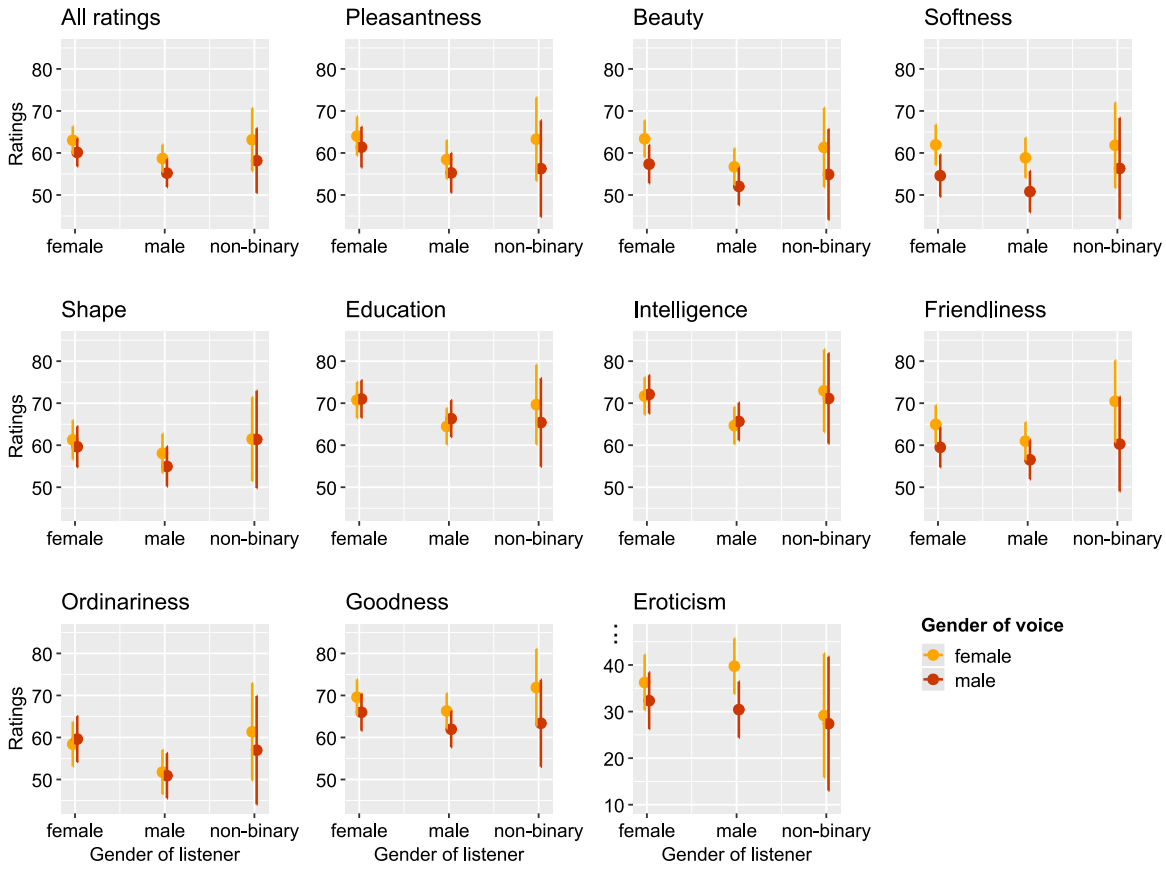


Figure 8: Overview of the interaction between GENDER OF LISTENER and GENDER OF VOICE in all models. Rating on the y-axis is always arranged so that higher values are more positive. Error bars indicate 95% CIs. Note that the range displayed is the same for all panels except for EROTICISM, which received lower ratings overall.

This pattern between men and women generally holds true across scales (Figure 8), and the comparisons are almost always significant (see Table 5 for an overview of significances; note that the sample size for non-binary listeners was substantially smaller). We can generalize that male voices are perceived by female and male listeners as comparatively negative, ugly, hard, unfriendly, evil, and unerotic. At the same time, men in general rate language or its speakers as less positive, less pleasant, uglier, less educated, more stupid, and stranger than women do.

There are, however, some cases where the data either augment or deviate from this pattern. One special case is EROTICISM, displaying a more extreme contrast between voices for male listeners (Figure 7, top right panel). Female, male, and non-binary listeners do not differ in how erotic they perceive either female or male voices. However, both female and male listeners highly significantly rate female voices as more erotic than male voices, and this effect is particularly strong for male listeners. This makes sense if we assume that most of our listeners are heterosexual. In contrast, the status dimension, represented by INTELLIGENCE and EDUCATION, encompasses cases that deviate from the general pattern. When we look at INTELLIGENCE (Figure 7, bottom left panel), we again find that male listeners rate both female and male voices as less intelligent than female listeners. However, the effect *between* voices disappears. Female voices are, in fact, not rated as more intelligent than male voices. This is true for ratings by all three listener genders. We should not overinterpret a null result, but it is indeed a conspicuously absent effect in the presence of the others. The picture is similar for the other of the two status scales, EDUCATION (Figure 7, bottom left panel).

To summarize the results thus far, Table 5 provides an overview of significant effects in all models. We can see that whether the language contains /x/ only makes a difference if we look at listeners who were exposed to that sound, and even then, CONDITION does not reach significance for individual scales. Meanwhile, RECOGNITION, FAMILIARITY, and GENDER make a notable difference in many cases. I will come back to these findings in the discussion in Section 4. The results already hint at the fact that for /x/, in this experiment, social factors may outweigh the simple distinction between stimulus and control. However, to examine how the indexical fares against the iconic, it can be useful to further quantify this comparison.

Figure 9 plots for each model how important the predictors are compared to each other in predicting, or splitting, the data, based on the increase in node purity in 500-tree random forests (see explanation in Section 2.6). The more node purity increases, the more important the predictor is. In each panel, the predictors are sorted by importance, from most important at the top to least important at the bottom. For instance, looking at the across-scale model in the top left panel, we see that the two central social predictors, FAMILIARITY and RECOGNITION, are among the most important predictors. Meanwhile, CONDITION, the phonological predictor, ranks rather low in comparison. If we compare all models for the different scales, we can see that the social predictors always stay on top, while the sound-based predictor always stays close to the bottom. We can use this information as another piece of the puzzle helping us to gauge the effect of the phonological makeup of a language with /x/ compared to the effects of social predictors on ratings.

Table 6: Overview of significant effects in all linear mixed-effects regression models, with interactions and EXPOSED^{UNCLEAR}, LANGUAGE, INPUT, OUTPUT, and LOCATION omitted. Lighter shaded cells represent negative effects, darker shaded cells represent positive effects. For terms included as interactions, [brackets] represent the reference level for which the comparison is true, e.g., CONDITION$^{ST\ [EX]}$ contrasts stimulus (against control) in the exposed group. Details for all models are documented in the OSF Repository.

Predictor	Rating	Pleasantness	Beauty	Softness	Shape	Education	Intelligence	Friendliness	Ordinariness	Goodness	Eroticism
CONDITION$^{ST\ [EX]}$	*										
CONDITION$^{ST\ [LE]}$											
EXPOSURE$^{EX\ [ST]}$	**		**			*	*		.	*	*
EXPOSURE$^{EX\ [CT]}$			.			*	*				.
RECOGNITION^{AS}											
RECOGNITION^{GE}	***				*						
RECOGNITION^{ME}	***				.			**		***	
RECOGNITION^{RO}	***	**			.			.		.	
RECOGNITION^{O}	***	**	*		*			.		**	
FAMILIARITY	***	***	***	***	***	***	***	***	***	***	***
GENDERV$^{M\ [LM]}$	***	*	***	***	*			***		***	***
GENDERV$^{M\ [LF]}$	***		***	***				***		**	**
GENDERL$^{M\ [VM]}$	***	**	**		.	*	***		***	.	
GENDERL$^{M\ [VF]}$	***	***	***			***	***	*	***		
POLYGLOT	.			.		.	.			.	.
AGE											
MUSICALITY		*									*
LINGUISTICS											

Abbreviations:

ST stimulus, **CT** control, **EX** exposed, **LE** less exposed, **AS** Asia, **GE** Germanic, **ME** Middle East, **RO** Romance, **O** other, **M** male, **LM** listener male, **LF** listener female, **VM** voice male, **VM** voice female, **GENDERV** = GENDER OF VOICE, **GENDERL** = GENDER OF LISTENER.

Significance codes: *** for p < .001, ** for p < .01, * for p < .05, . for p < .1, no asterisks for p > .05. Tukey adjustment of p-values applied to post-hoc comparisons of multiple contrasts (e.g., CONDITION × EXPOSURE).

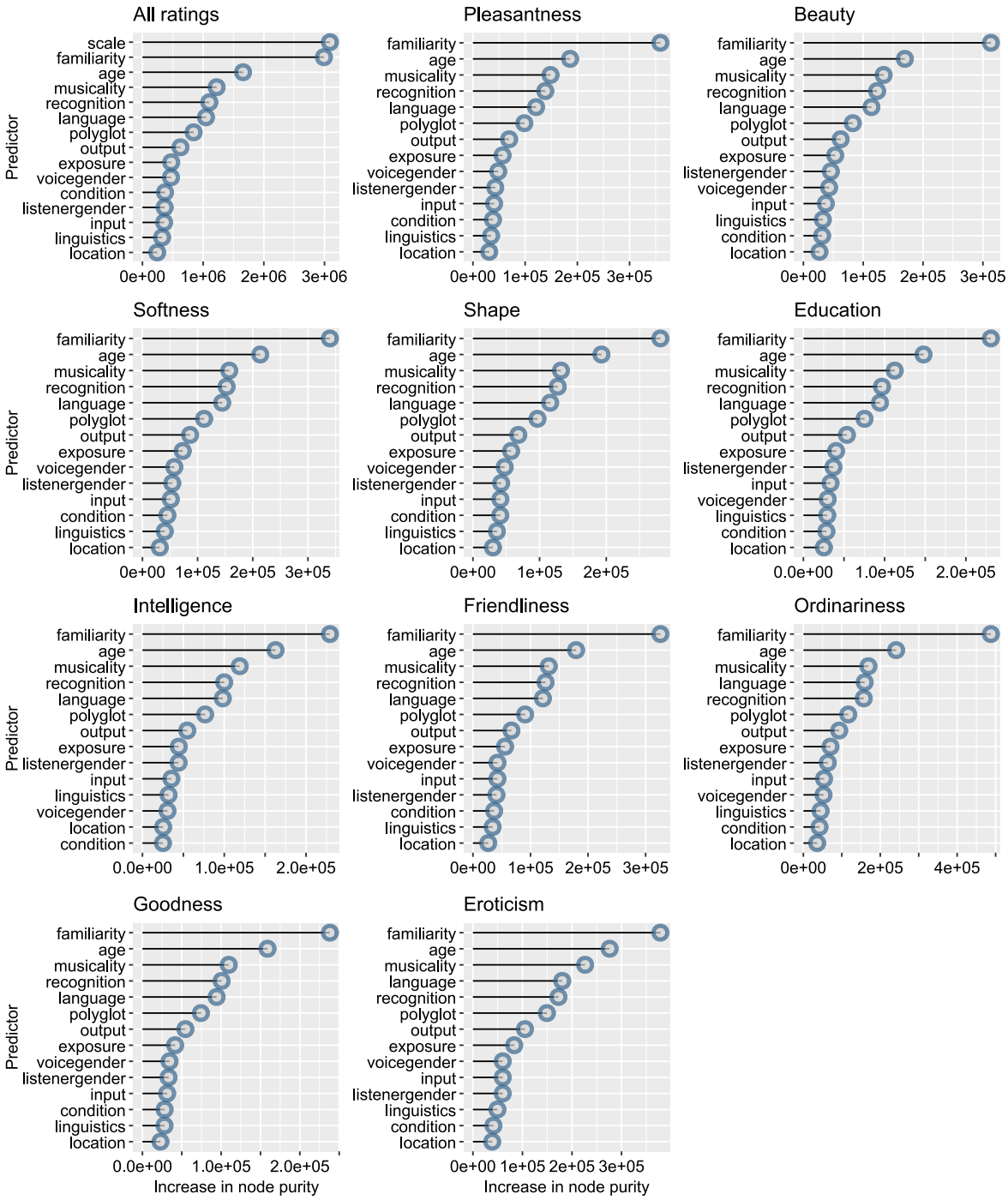


Figure 9: Overview of variable importance for all models in predicting the ratings, based on the increase in node purity in 500-tree random forests.

4. Discussion

This study investigated the interplay between indexicality and iconicity by bringing the two mechanisms together in a controlled study, using /x/ as a test case. Let us now discuss how we feel about /x/, and what this implies for social and aesthetic meaning-making.

4.1 Does /x/ sound inherently bad?

In this experiment, ratings were not, or only weakly, affected by whether /x/ is present, while several social variables turned out to be strong predictors. That is, we find support for $H1_{INDEXICALITY}$ and $H2_{INDEXICALITY}$ rather than $H1_{ICONICITY}$ and $H2_{ICONICITY}$. This study, then, is a case where indexical factors strongly contribute to evaluative meaning—perhaps more so than sound itself. This finding is in line with other studies that found social variables to outweigh phonetic-phonological variables (e.g., Reiterer et al., 2020). It is also consistent with the suggestion that iconicity can be overridden by indexicality. For example, it has been suggested that the frequency code, i.e., the idea that low frequency sounds are perceived as dominant because they are used by larger-sized animals (Ohala, 2010), may not apply when overridden by social meaning, such as when creaky voice is not perceived as authoritative or competent even though it is low frequency (Winter et al., 2021).

To some extent, of course, the comparison is unfair, as we would not necessarily expect a single predictor based on a single sound to overshadow all other predictors. However, while the presence of /x/ as such did not make a difference, in interactions with exposure stimuli with /x/ were rated worse if the listener had more likely been exposed to /x/. This is evidence that iconic effects are modulated by indexicality. The tentative explanation for the direction of this

effect is that exposed listeners are more aware of respective stereotypes, and consequently rate languages with /x/ worse despite being more used to them (but see Section 4.3). Indeed, the strongest predictors in the models were sociocultural, not phonetic-phonological, in nature. Listeners rate language worse if they are male, if they perceive the language as being less familiar, and if they *felt* the language sounded similar to a specific language group, such as those associated with "guttural sounds" or "harshness", like languages from the Middle East or Germanic languages. This is in line with pervasive stereotypes about how languages sound that belong to these categories. However, the Latin Lover effect could not be replicated. This could be due to the inherent lack of eroticism in robots, which served as speakers in the story. It could also be because the present study had a lower share of WEIRD participants than some earlier studies (Henrich et al. 2010), who may be more familiar with this stereotype. Finally, it is also worth considering that RECOGNITION is a notoriously mysterious predictor with unreliable answers. In the study by Anikin et al. (2023), which tested natural languages, listeners rated languages better when they *thought* they recognized them, although they were actually wrong half of the time. They also failed to report recognizing a language even when they performed the experiment in that language. The present study, testing constructed language, used a more complex operationalization based on open answers about which natlangs they resembled, but this does not preclude these issues. Recordings were rated best when listeners thought the language resembled languages from Africa or Asia. This may either be a "positive" exoticism effect, or, as several listeners lived in countries like South Africa or Japan, a familiarity effect. Future studies could try to decode the mapping between listener demographics and effects of social categorization in a more targeted way. In general, the fact that how listeners categorize language strongly determines evaluative meaning even if incorrect is a quintessentially indexical finding

and lends further support to the idea that "social categorisation mediates the language attitudes process" (Dragojevic & Goatley-Soan, 2020, p. 12; Ryan, 1983).

Taken together, the finding that, at least in the case of /x/, sociocultural predictors modulate (EXPOSURE) or overshadow (GENDER, FAMILIARITY, RECOGNITION) the influence of the phonetic-phonological makeup of language on attitudes lends further support to the indexical route (Figure 1; Peirce, 1958; Silverstein, 2003; Giles & Niedzielski, 1998). Even potentially iconic associations with certain types of sound might become manifest only through an indexical detour, through which listeners rely on previous experience with languages and groups of speakers. This detour may partly be enabled, or indeed required, by what Winter et al. (2019) call *pluripotentiality*: One phonological pattern can invoke different semantics. Only social context can make specific semantics become manifest—a quantum-superposed Schrödinger's /x/, so to speak, whose meaning settles only through social observation. A similar semantic potential is well-known in sociolinguistics. Listeners evaluate the same linguistic features in very different ways, not only because they categorize speakers differently (and often incorrectly), but also because language varieties index many different identities, any of which could emerge as salient in a given context (see Dragojevic & Goatley-Soan, 2022 for discussion). In a nutshell, whether we dislike sounds depends on who uses them.

Given the strong support for the indexical route, the present results also speak in favor of the imposed norm and social connotations hypotheses (Giles et al., 1979; Trudgill & Giles, 1976). But what, then, about the inherent value hypothesis? This hypothesis usually has strong theoretical backup due to its rationale being rooted in several statistical and biological patterns. As discussed by Barker and Bozic (2024), some associations with sound could be universal due to, for instance, co-occurrence or experience (e.g., smaller and spikier things produce higher

693 frequencies when hitting a surface), evolved properties (e.g., small animals produce high-fre-
694 quency sounds), physical resemblance (e.g., lip rounding for roundness), or embodiment and
695 imitation (speech could have evolved from proto-dialogue based on iconic hand gestures, sup-
696 ported by the mirror neuron system, by which iconicity bootstraps language acquisition). As
697 discussed in Section 1, in the case of /x/, evolved properties in particular (e.g., coughing; threat-
698 ening animal sounds, Rendall & Owren, 2010) and their shared resemblances with /x/ lend
699 plausibility to the idea that this sound may be iconically associated with illness, choking, growl-
700 ing, and hostility (Mooshammer et al., 2023; Stanley, 2003). But this experiment barely found
701 support for stimuli with /x/ being rated worse than the control condition without /x/, unless it
702 looked at listeners who had been exposed to it. Even then, this effect was significant only for
703 pleasantness, shape, and across scales. How, then, can we reconcile these ideas of evolutionary
704 and acquisition-inspired mechanisms with the results of the present study?

705 4.2 A fundamentally indexical system?

706 One of the most important insights of the present study is that indexical effects emerge even if
707 almost all social context is removed. In other words, it is possible to observe social-indexical
708 variation we know from classic sociolinguistics even though the indexical route is intentionally
709 weakened by removing the stimuli from real life context. This is in line with Li and Roberts
710 (2023), who showed that both first-order indexicality (i.e., the connection between sound and
711 social groups; Silverstein, 2003) and higher-order indexicality (e.g., the connection between
712 social groups and perceived social attributes) can emerge in constructed language. This indi-
713 cates that indexicality might be a fundamental property of meaning-making. As such an

integral property, indexicality might constrain and even encompass iconicity in various ways, which may help explain the apparent discrepancy between the present findings and iconicity as a strong theoretical and empirical notion. Indeed, some kinds of what is sometimes subsumed under the label "iconicity" are much closer in their mechanism to indexicality than previously thought.

First, some kinds of iconicity do rely on learned associations rather than evolutionary hardwiring. For example, systematicity (Dingemanse et al., 2015) refers to sound-meaning mappings that are considered diagrammatically iconic, meaning the sound is systematically linked to its meaning without actually resembling it. For instance, phonaesthemes like *gl-* in English do not naturally evoke the meaning of 'light'; rather this meaning is acquired from a pattern existing in the English language. As Haslett and Cai (2023, p. 630) write that here, "systematicity produces iconicity, no resemblance required". An iconic word like *meow* 'the sound of a cat' may directly imitate the sound of a cat, but an iconic sound like /i/ 'smallness' or /a/ 'bigness' must first come to be associated with size, similar to a social-indexical association like *pen* [pɪn] 'uneducated'. It is a learned association (learned perhaps via mouth shape, or via typical frequency patterns in small and large animals, Ohala, 2010), which goes beyond onomatopoeia. Here, iconicity can still seem to play a role on the surface even when the underlying mechanism by which the sound-meaning mapping comes about is rather indexical. Future studies could attempt to investigate whether this is the case for /x/.

Second, it has been suggested that indexicality can lead to iconicity, that indexical features can *become* iconic (cf. the ideas of iconization and rhematization, Irvine & Gal, 2000 and Irvine, 2022; or the idea of indexical iconicity, Silverstein, 2003). Linguistic features may appear to inherently encode social meaning when their indexical link has become naturalized enough

(Hall, 1999, p. 511). In other words, certain characteristics of social groups are perceived as being inherent in the linguistic features with which they are associated. For example, a realization of *pen* as [pɪn] may be classified as inherently 'ignorant' without pointing to Southern US speakers for the listener (Preston, 2010). This echoes findings from qualitative data showing listeners to retroactively rationalize a connection between sound and meaning as being iconic, like "You can almost hear Scottish people mining in a shaft when they talk" (Stein, 2023). Evaluative meaning has become so naturalized within the sound that the listener can presumably skip first-order indexicality and directly access attitudes without social categorization. This could partly explain why exposed listeners rated /x/ worse, namely, because the negative perception of the people using that sound has become engrained in it.

Finally, there is an overlap of indexicality and iconicity in that both, contrary to received wisdom, feature elements of arbitrariness. In the traditional tripartite distinction, *iconic* signs have (some of) the characteristics of the signified for which they act as signifiers, *indexical* signs have some real-life connection to the objects they represent but no inherent resemblance or shared quality, while only the relation of *symbolic* signs to their objects is purely arbitrary and conventional (Peirce, 1958; Sebeok, 2001; Silverstein, 2003). Consequently, some researchers believe that iconicity competes with arbitrariness, or that iconicity is defined as non-arbitrariness (Barker & Bozic, 2024). However, contrary to this understanding (or, some would say, misrepresentation of Peircean semiotics), some researchers emphasize that iconicity can, or must, coexist with arbitrariness (e.g., Cohn & Schilperoord, 2024; Anderson, 1998, p. 29; Perniss & Vigliocco, 2014). An example of this is onomatopoeia, which is also partly conventionalized and acquired, and differs between languages (Barker & Bozic, 2024; Kwon, 2016; Occhino et al., 2017; Nielsen & Dingemanse, 2021). It combines iconicity, indexicality, and

symbolicity, as the onomatopoetic expression is an iconic image that points to the original sound but is restricted by the phonemic inventory (Körtvélyessy & Štekauer, 2024). Iconicity is thus arbitrary in the sense that a given iconic association is but one of many associations listeners could have in a complex, pluripotential (Winter et al., 2019) system of many-to-many mappings of meaning to a limited sound system. Indexicality, meanwhile, is also arbitrary in the sense that the connection between social groups and perceived social attributes, or the second-order sign, is, at best, stereotypical and thus conventionally agreed upon. Even the connection between sound and social groups is a product of history and culture, rather than something that is naturally given or even causally related (such as in the classic example of an index in which smoke points to fire).

Systematicity, iconization, and arbitrariness are three examples that expose the theoretical difficulties researchers face in attempting to clearly separate indexicality and iconicity. The strong support for indexical effects found in the present study can thus be reconciled with the idea of iconicity as a fundamental property of human cognition by acknowledging that in many cases, indexicality may actually masquerade as iconicity. Indexicality and iconicity need not be opposed. Both can contribute to attitudes in conspiracy, combining phonology, synaesthesia, and culture (Berthele, 2010). Reiterer et al. (2020) argue that this conspiracy could be similar to the interplay of culture and biology in the aesthetics of music. The strong effects both these studies and the present experiment found for familiarity, in particular, could mirror musical anticipation effects (Kogan & Reiterer, 2021). Even arguments from the ontogenesis and phylogenesis of language do not have to rely on iconicity, but can just as well be made for indexicality. It has often been suggested that learning associations between sound and meaning, or cue and outcome, may be easier with features that are socially salient (Li & Roberts, 2023; Rácz

et al., 2020). This implies that indexicality, too, would be an important ontogenetic and phylogenetic bootstrap mechanism. Both iconically (Anikin et al., 2023) and indexically (Kristiansen, 2011) motivated preferences could shape the typological frequency of linguistic features in the long run.

4.3 Where do we go from here?

I believe the present study, including the SSPG and its overall experimental design, provides a useful template for further studies and constitutes a promising step towards investigating indexical and iconic effects in language attitudes for many different sounds and other features of language. There are a few limitations of the present study that future research should keep in mind.

First, one obvious possibility is that /x/ may not have been salient enough to produce strong effects. Given the experimental setup, however, I consider this unlikely. About 20 of 100 words in each stimulus contained /x/, and as discussed in Section 2.3, much research shows that listener evaluations drastically change based on even just one single token of a phonological marker. Of course, this experiment removed as much real-life social context as possible, using constructed language and a fictional context. Listeners learn indexical associations likely due to the relative salience and social value of a given trait in a given context (Li & Roberts, 2023; Rácz et al., 2020). Social salience even modulates speech perception itself, so that listeners categorize the same sound as different phonemes (e.g., TRAP or LOT) depending on its connection to different social personas (D'Onofrio, 2018). This means that by taking real-life context

away we intentionally weaken indexical effects. It is even more striking, then, that social variables turned out to be strong predictors of rating in the present study.

Second, this experiment is also subject to the usual limitations of rating studies, most prominently the social desirability bias. Participants are hesitant to expose their racism, classism, and of course linguicism (see, e.g., Mooshammer et al., 2023, p. 24). This study tried to mitigate this bias by using a story about unknown robots from space. For rating synthetic language, researchers face the additional challenge that phonetic realizations always vary with language optimization (as explained in Section 2.1). This experiment embraced this reality consciously by using multiple language optimizations. But that also means that there was variation between languages. For example, Dutch voices would not realize /w/ as [w], which accounting for borrowings exists in Dutch, but they would consistently produce [ʊ], even when forcing Polly to realize the sounds as intended with IPA transcription in SSML tags. An individual participant always listened to stimulus-control contrasts within one voice and within one language optimization (see Section 2.3), which meant the experiment was able to include language optimization as a covariate in the statistical modeling (see Section 2.5) as a strategy to mitigate against this limitation. Manual checks of the sound files also revealed that realizations of /x/ as [ç] instead of [x], which could have been expected due to allophonic variation in some of the optimizations' languages, did not turn out to be a systematic error (with very few exceptions for only some tokens preceding /i/ in some German and Dutch voices). Still, it could be interesting for future studies to investigate differences in ratings between languages that differ in their realizations of particular phonemes. It is also conceivable that ratings may differ depending on the phonological context in which the segment of interest occurs. For example, German speakers may rate /x/ differently when it occurs in word-initial position, which it does not in

826 German. Future studies could introduce different phonotactic rules to their stimuli, or restrict
827 certain segments to certain positions in different syllable structures, to test such hypotheses.[2]

828 Third, future studies are needed to disentangle the effects of exposure and familiarity.
829 Exposed listeners rated /x/ worse (not better), which I suggested could be interpreted as inter-
830 nalized stereotyping. However, at first glance, this finding seems to be at odds with the famili-
831 arity effect, which shows that listeners who indicated they were more familiar with the con-
832 structed language rated it better. One potential explanation for this supposed inconsistency is
833 that the exposure predictor may differ in nature from the familiarity predictor. Self-reported
834 familiarity (i.e., what people *feel* sounds "familiar") may tap into a different underlying mech-
835 anism than exposure measured by listener language experience (i.e., a more objective measure
836 of the sounds listeners actually encounter in their daily lives). Both linguistic and psychological
837 studies may find it worthwhile to try resolving this discrepancy between measured and per-
838 ceived familiarity.

839 Finally, one major challenge for studies of this type, beyond disentangling different in-
840 dexical effects, is to decide on which measures are valid measures of indexicality in the first
841 place, and how they are to be operationalized and implemented statistically. As a reviewer
842 points out, we would not like to assume that any non-iconic effect that cannot be traced back
843 to a manipulation of the signal itself is automatically indexical in nature. At the same time, the

[2] Following the suggestion of a reviewer, I also tested for the influence of phonotactic restrictions by further differentiating different types of exposure. I recoded EXPOSURE to a four-level factor: `exposed` (/x/ occurs both in syllable-final and -initial position, i.e., Arabic, Spanish, and Polish), `less exposed in initial position` (/x/ does not or only rarely occur in initial position, i.e., German and Dutch), least `exposed` (/x/ does not occur at all or is extremely rare in any position, i.e., English, Italian, and Japanese), and `unclear`. The results show that the category `less exposed` in initial position patterns in between exposed and least exposed participants, for both stimulus and control ratings. The intermediate steps from less exposed in initial position to exposed, and to least exposed, respectively, do not reach statistical significance; neither does the difference within the condition `less exposed in initial position` between stimulus and control. The interested reader can access these additional analyses in the supplementary materials.

predictors that were selected for this study to represent indexical processes can strongly be argued to tap into indexicality indeed: As outlined in Section 1, the iconic route assumes that associations arise due to a shared resemblance between form and meaning, which is why iconic effects are often hypothesized (with some support, see, e.g., the evidence from bouba-kiki studies) to be largely independent of environmental background. Consequently, from the iconic perspective we would expect ratings of the stimulus to significantly differ regardless of participant background. However, from the indexical perspective participant background is expected to matter such that what affects the perception of linguistic features is experience. Regardless of linguistic properties, experience (listeners' upbringing, socialization, exposure to sociocultural meaning; in a sense, nurture rather than nature) is expected to shape how we feel about language. This is why the indexical route in the present study is operationalized through variables like EXPOSURE, RECOGNITION, FAMILIARITY, and the sociocultural background of the listener (e.g., GENDER).

A challenge of statistical implementation of these indexical measures is the possibility that these indexical predictors interact with the manipulation (see discussion in Section 4.2 on the theoretical challenges of keeping iconicity and indexicality apart) and could thus be included as interactions. In the present study, only EXPOSURE was allowed to interact with CONDITION, since EXPOSURE is coded to be informative about just one of the two conditions (exposure to /x/). Other interactions were not expected; the indexical perspective would not necessarily predict effects to differ based on the signal. For example, we would expect listeners to rate the stimulus condition better when they think of it as a Romance language and worse when they think of it as a Germanic language, but we would also expect them to rate the control condition better when they perceive it as Romance, and the control condition worse if they

perceive it as Germanic. However, different predictors could still interact with each other regardless of the hypotheses derived from indexicality itself. Additional analyses with new models including interactions of CONDITION with LANGUAGE, GENDER OF VOICE, RECOGNITION, and FAMILIARITY, respectively, confirm many of the already known effects (e.g., that male voices are rated worse for both stimulus and control) but show that none of the interactions are statistically significant with the exception of CONDITION × FAMILIARITY: Here, both the stimulus and the control are rated better with increasing self-reported familiarity, but the stimulus a little less so (i.e., the effect is a little less strong for the language with /x/). Since the difference is modest, we should not overinterpret the finding, but we may very cautiously read this as weak support for the iconic route shining through. If this were true, it could mean that when we feel a language sounds familiar, the negative effect of /x/ itself inhibits how positively this makes us feel about that language. The interested reader can access these additional analyses in the OSF Repository.

A challenge of validity for the indexical measures, at least for EXPOSURE, is that there also exists a more general exposure effect of the priming type by which listeners rate stimuli better the more often they hear them (referred to as "mere exposure" in the literature, e.g., Zajonc, 1968). In the present study, mere exposure is not a likely explanation for the effects found, as more exposed listeners rated the stimulus condition worse, rather than better, than the control. More fundamentally, however, it is a matter of theoretical belief whether mere exposure should still be considered to tap into socialization, or if, indeed, exposure to language can ever be devoid of social meaning. Most sociolinguists would likely disagree with this, and I too tend to do so, since exposure, especially as operationalized in the present study (occurrence of the cue /x/ in the listeners' primary language), does not happen in isolation or without context, i.e., it is

not "mere". Being exposed to a linguistic feature changes how we perceive that feature socially—even repeating stimuli to someone in a lab could be regarded a form of social conditioning. Exposure to forms and what they mean to humans is in essence an indexical process. Whatever the case may be, this theoretical point contains an important lesson, which is that studies like the present one should make as transparent as possible their different operationalizations of indexicality and iconicity and the interpretations of measures that are supposed to reflect it. This way, the reader can make a better-informed decision when contextualizing the results, depending on where they stand theoretically.

Taken together, this line of research could support a "cognitive turn" in attitude research (Berthele, 2010) that enables us to re-evaluate some of the fundamentals of received linguistic wisdom, and refurbish its semiotic legacy of both indexicality and iconicity research.

5. Conclusion

Sounds may not only distinguish meaning, they *have* meaning. But what kind of meaning? And how does evaluative meaning, and hence language attitudes, emerge? This study explored this question with /x/ as a test case. Put plainly, it found that /x/ may not be as bad as people think it is—unless sociocultural factors *make* it sound bad. This supports the idea of iconicity being embedded into and constrained by a deeply indexical system.

The study provides a foundation based on which other linguistic features can be tested. This line of research will provide new insights into the interplay of iconicity and indexicality, and into the link between sound and meaning. It will also have implications of interest outside the academic community for the linguistically informed fight against linguicism. The hope is

that being more informed about why people feel the way they feel about language will, ultimately, help us combat linguicism and linguistic discrimination.

Data availability

The experimental materials, data, and code are available in the OSF Repository.

Acknowledgements

This research is funded by *redacted for review* — Project number *redacted for review*.

Ethics statement

This research was approved by the Ethics Commission of the Faculty of Arts and Humanities for non-invasive research on humans at *redacted for review*.

Competing interests

The author(s) declare none.

References

Anderson, E. R. (1998). *A grammar of iconism*. Fairleigh Dickinson University Press.

Anikin, A., Aseyev, N., & Johansson, N. E. (2023). Do some languages sound more beautiful than others? *Proceedings of the National Academy of Sciences of the United States of America*, *120*(17), e2218367120. https://doi.org/10.1073/pnas.2218367120

927 Baggia, P., Bagshaw, P., Bodell, M., Huang, D. Z., Xiaoyan, L., McGlashan, S., Tao, J., Jun, Y., Fang, H., Kang, Y.,
928 Meng, H., Xia, W., Hairong, X., & Wu, Z. (2010). *Speech Synthesis Markup Language (SSML) Version 1.1*.
929 Version https://www.w3.org/TR/speech-synthesis/

930 Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, *17*(8), 645–648.
931 https://doi.org/10.1111/j.1467-9280.2006.01759.x

932 Barker, H., & Bozic, M. (2024). The forms, mechanisms, and roles of iconicity: A review. *Advance Preprints*.
933 https://doi.org/10.31124/advance.171233725.51126728/v1

934 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of*
935 *Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

936 Bayard, D., Weatherall, A., Gallois, C., & Pittam, J. (2001). Pax Americana? Accent attitudinal evaluations in New
937 Zealand, Australia and America. *Journal of Sociolinguistics*, *5*(1), 22–49. https://doi.org/10.1111/1467-
938 9481.00136

939 Bertamini, M., Palumbo, L., Gheorghes, T. N., & Galatsidas, M. (2015). Do observers like curvature or do they
940 dislike angularity? *British Journal of Psychology*, *107*(1), 154–178. https://doi.org/10.1111/bjop.12132

941 Berthele, R. (2010). Investigations into the folk's mental models of linguistic varieties. In D. Geeraerts, G.
942 Kristiansen, & Y. Peirsman (Eds.), *Advances in Cognitive Sociolinguistics* (pp. 265–290). De Gruyter Mouton.
943 https://doi.org/10.1515/9783110226461.265

944 Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound-meaning
945 association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences of*
946 *the United States of America*, *113*(39), 10818–10823. https://doi.org/10.1073/pnas.1605782113

947 Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness
948 perception. *Animal Behaviour*, *82*(1), 55–59. https://doi.org/10.1016/j.anbehav.2011.03.024

949 Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological*
950 *Bulletin*, *106*(2), 265–289. https://doi.org/10.1037/0033-2909.106.2.265

951 Calhoun, S., Warren, P., Mills, J., & Agnew, J. (2024). Socialising the Frequency Code: Effects of gender and age
952 on iconic associations of pitch. *The Journal of the Acoustical Society of America*, *156*(5), 3183–3203.
953 https://doi.org/10.1121/10.0034354

954 Clopper, C. G., & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of American English
955 regional dialects. *Journal of Phonetics*, *32*(1), 111–140. https://doi.org/10.1016/s0095-4470(03)00009-3

956 Cohn, N., & Schilperoord, J. (2024). *A multimodal language faculty*. Bloomsbury Academic.
957 https://doi.org/10.5040/9781350404861

958 Coupland, N., & Bishop, H. (2007). Ideologised values for British accents. *Journal of Sociolinguistics*, *11*(1), 74–93.
959 https://doi.org/10.1111/j.1467-9841.2007.00311.x

960 D'Onofrio, A. (2018). Personae and phonetic detail in sociolinguistic signs. *Language in Society*, *47*(4), 513–539.
961 https://doi.org/10.1017/s0047404518000581

962 Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity,
963 and systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615.
964 https://doi.org/10.1016/j.tics.2015.07.013

965 Domizi, A. (2024). *Die ästhetische Wahrnehmung der deutschen Sprache im europäischen Raum*. IDS-Verlag.
966 https://doi.org/10.14618/amades-62

967 Dragojevic, M., & Goatley-Soan, S. (2020). Americans' attitudes toward foreign accents: Evaluative hierarchies
968 and underlying processes. *Journal of Multilingual and Multicultural Development*, *43*(2), 167–181.
969 https://doi.org/10.1080/01434632.2020.1735402

970 Dragojevic, M., & Goatley-Soan, S. (2022). The verbal-guise technique. In R. Kircher & L. Zipp (Eds.), *Research*
971 *methods in language attitudes* (pp. 203–218). Cambridge University Press.
972 https://doi.org/10.1017/9781108867788.017

973 Essick, G. K., James, A., & McGlone, F. P. (1999). Psychophysical assessment of the affective components of non-
974 painful touch. *NeuroReport*, *10*(10), 2083–2087. https://doi.org/10.1097/00001756-199907130-00017

975 Essick, G. K., McGlone, F., Dancer, C., Fabricant, D., Ragin, Y., Phillips, N., Jones, T., & Guest, S. (2010).
976 Quantitative assessment of pleasant touch. *Neuroscience & Biobehavioral Reviews*, *34*(2), 192–203.
977 https://doi.org/10.1016/j.neubiorev.2009.02.003

978 Fasoli, F., & Maass, A. (2020). The social costs of sounding gay: Voice-based impressions of adoption applicants.
979 *Journal of Language and Social Psychology*, *39*(1), 112–131. https://doi.org/10.1177/0261927x19883907

980 Giles, H., Bourhis, R., & Davies, A. (1979). Prestige speech styles: The imposed norm and inherent value
981 hypotheses. In W. C. McCormack & S. A. Wurm (Eds.), *Language and society* (pp. 589–596). De Gruyter
982 Mouton. https://doi.org/10.1515/9783110806489.589

983 Giles, H., Bourhis, R., Lewis, A., & Trudgill, P. (1974). The imposed norm hypothesis: A validation. *Quarterly*
984 *Journal of Speech*, *60*(4), 405–410. https://doi.org/10.1080/00335637409383249

985 Giles, H., & Johnson, P. (1987). Ethnolinguistic identity theory: A social psychological approach to language
986 maintenance. *ijsl*, *1987*(68), 69–100. https://doi.org/10.1515/ijsl.1987.68.69

987 Giles, H., & Niedzielski, N. (1998). Italian is beautiful, German is ugly. In L. Bauer & P. Trudgill (Eds.), *Language*
988 *myths* (pp. 85–93). Penguin.

989 Hall, S. (1999). Encoding, decoding. In S. During (Ed.), *The cultural studies reader* (2 ed., pp. 507–517). Routledge.

990 Haslett, D. A., & Cai, Z. G. (2023). Systematic mappings of sound to meaning: A theoretical review. *Psychonomic*
991 *Bulletin & Review*, *31*(2), 627–648. https://doi.org/10.3758/s13423-023-02395-y

992 Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*(1), 575–604.
993 https://doi.org/10.1146/annurev.psych.53.100901.135109

994 Hilton, N. H., Gooskens, C., Schüppert, A., & Tang, C. (2022). Is Swedish more beautiful than Danish? Matched
995 guise investigations with unknown languages. *Nordic Journal of Linguistics*, *45*(1), 30–48.
996 https://doi.org/10.1017/s0332586521000068

997 Irvine, J. T. (2022). Revisiting theory and method in language ideology research. *Journal of Linguistic Anthropology*,
998 *32*(1), 222–236. https://doi.org/10.1111/jola.12335

999 Irvine, J. T., & Gal, S. (2000). Language ideology and linguistic differentiation. In P. V. Kroskrity (Ed.), *Regimes of*
1000 *language* (pp. 35–83). School of American Research Press.

1001 Kawahara, S., Godoy, M. C., & Kumagai, G. (2021). English speakers can infer Pokémon types based on sound
1002 symbolism. *Frontiers in Psychology*, *12*, 648948. https://doi.org/10.3389/fpsyg.2021.648948

1003 Kircher, R., & Zipp, L. (2022). An introduction to language attitudes research. In R. Kircher & L. Zipp (Eds.),
1004 *Research methods in language attitudes* (pp. 1–16). Cambridge. https://doi.org/10.1017/9781108867788.002

1005 Kogan, V. V., & Reiterer, S. M. (2021). Eros, beauty, and phon-aesthetic judgements of language sound. We like
1006 it flat and fast, but not melodious. Comparing phonetic and acoustic features of 16 European languages.
1007 *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.578594

1008 Köhler, W. (1929). *Gestalt psychology*. Horace Liveright.

1009 Körner, A., & Rummer, R. (2022). Articulation contributes to valence sound symbolism. *Journal of Experimental*
1010 *Psychology: General*, *151*(5), 1107–1114. https://doi.org/10.1037/xge0001124

1011 Körtvélyessy, L., & Štekauer, P. (2024). Introduction: Why onomatopoeia? In L. Körtvélyessy & P. Štekauer (Eds.),
1012 *Onomatopoeia in the world's languages*. De Gruyter Mouton. https://doi.org/10.1515/9783111053226-001

1013 Kristiansen, T. (2011). Attitudes, ideology and awareness. In R. Wodak, B. Johnstone, & P. Kerswill (Eds.), *The*
1014 *SAGE handbook of sociolinguistics*. SAGE. https://doi.org/10.4135/9781446200957

1015 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). *lmerTest: Tests in linear fixed effects models*.
1016 Version 3.1.2. https://cran.r-project.org/web/packages/lmerTest/index.html

1017 Kwon, N. (2016). Empirically observed iconicity levels of English phonaesthemes. *Public Journal of Semiotics*, *7*(2),
1018 73–93. https://doi.org/10.37693/pjos.2016.7.16470

1019 Ladegaard, H. J. (1998). National stereotypes and language attitudes: The perception of British, American and
1020 Australian language and culture in Denmark. *Language & Communication*, *18*(4), 251–274.
1021 https://doi.org/10.1016/s0271-5309(98)00008-1

1022 Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken
1023 languages. *Journal of Abnormal and Social Psychology*, *60*, 44–51. https://doi.org/10.1037/h0044430

1024 Lattner, S., Meyer, M. E., & Friederici, A. D. (2004). Voice perception: Sex, pitch, and the right hemisphere. *Human*
1025 *Brain Mapping*, *24*(1), 11–20. https://doi.org/10.1002/hbm.20065

1026 Leiner, D. J. (2024). *SoSci Survey*. Version 3.5.01. https://www.soscisurvey.de

1027 Lenth, R. V., & Piaskowski, J. (2026). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. Version
1028 2.0.3. https://doi.org/10.32614/CRAN.package.emmeans

1029 Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility.
1030 *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

1031 Lev-Ari, S., & McKay, R. (2022). The sound of swearing: Are there universal patterns in profanity? *Psychonomic*
1032 *Bulletin & Review*. https://doi.org/10.3758/s13423-022-02202-0

1033 Li, A., & Roberts, G. (2023). Co-occurrence, extension, and social salience: The emergence of indexicality in an
1034 artificial language. *Cognitive Science*, *47*(5), e13290. https://doi.org/10.1111/cogs.13290

1035 Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.
1036 https://CRAN.R-project.org/doc/Rnews/

1037 Lopes, D. M. (2017). Disputing taste. In J. O. Young (Ed.), *The semantics of aesthetic judgements* (pp. 61–81).
1038 Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198714590.003.0004

1039 Marks, L. E. (1975). On colored-hearing synesthesia: Cross-modal translations of sensory dimensions.
1040 *Psychological Bulletin*, *82*(3), 303–331. https://doi.org/10.1037/0033-2909.82.3.303

1041 Maschmann, I. T., Körner, A., Boecker, L., & Topolinski, S. (2020). Front in the mouth, front in the word: The
1042 driving mechanisms of the in-out effect. *Journal of Personality and Social Psychology*, *119*(4), 792–807.
1043 https://doi.org/10.1037/pspa0000196

1044 Menn, L., & Vihman, M. (2011). Features in child phonology: Inherent, emergent, or artefacts of analysis? In G.
1045 N. Clements & R. Ridouane (Eds.), *Where do phonological features come from? Cognitive, physical and*
1046 *developmental bases of distinctive speech categories* (pp. 261–302). John Benjamins.
1047 https://doi.org/10.1075/lfab.6.10men

1048 Mooshammer, C., Bobeck, D., Hornecker, H., Meinhardt, K., Olina, O., Walch, M. C., & Xia, Q. (2023). Does
1049 Orkish sound evil? Perception of fantasy languages and their phonetic and phonological characteristics.
1050 *Language and Speech*, 238309231202944. https://doi.org/10.1177/00238309231202944

1051 Munson, B., & Babel, M. (2007). Loose lips and silver tongues, or, projecting sexual orientation through speech.
1052 *Language and Linguistics Compass*, *1*(5), 416–449. https://doi.org/10.1111/j.1749-818x.2007.00028.x

1053 Nielsen, A. K. S., & Dingemanse, M. (2021). Iconicity in word learning and beyond: A critical review. *Language*
1054 *and Speech*, *64*(1), 52–72. https://doi.org/10.1177/0023830920914339

O'Connor, J. J. M., & Barclay, P. (2017). The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior*, *38*(4), 506–512. https://doi.org/10.1016/j.evolhumbehav.2017.03.001

Occhino, C., Anible, B., Wilkinson, E., & Morford, J. P. (2017). Iconicity is in the eye of the beholder. *Gesture*, *16*(1), 100–126. https://doi.org/10.1075/gest.16.1.04occ

Ohala, J. J. (2010). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 325–347). Cambridge University Press. https://doi.org/10.1017/cbo9780511751806.022

Palumbo, L., Ruta, N., & Bertamini, M. (2015). Comparing angular and curved shapes in terms of implicit associations and approach/avoidance responses. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0140043

Peirce, C. S. (1958). *Selected writings*. Dover.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130300. https://doi.org/10.1098/rstb.2013.0300

Pisanski, K., & Feinberg, D. R. (2018). Vocal attractiveness. In S. Frühholz, P. Belin, K. Pisanski, & D. R. Feinberg (Eds.), *The Oxford handbook of voice perception* (pp. 606–626). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.27

Preston, D. R. (2010). Language, people, salience, space: Perceptual dialectology and language regard. *Dialectologia*, *5*, 87–131.

Preston, D. R. (2017). Perceptual dialectology. In C. Boberg, J. Nerbonne, & D. Watt (Eds.), *The handbook of dialectology* (pp. 177–203). Wiley. https://doi.org/10.1002/9781118827628.ch10

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, *18*(1), 10–30. https://doi.org/10.1177/0261927x99018001002

R Core Team. (2025). *R: A language and environment for statistical computing*. Version 4.5.2. R Foundation for Statistical Computing. http://www.R-project.org/

Rácz, P., Hay, J., & Pierrehumbert, J. B. (2020). Not all indexical cues are equal: Differential sensitivity to dimensions of indexical meaning in an artificial language. *Language Learning*, *70*(3), 848–885. https://doi.org/10.1111/lang.12402

Ramachandran, V. S., & Hubbard, E. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34.

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. https://doi.org/10.1207/s15327957pspr0804_3

Reiterer, S. M., Kogan, V., Seither-Preisler, A., & Pesek, G. (2020). Foreign language learning motivation: Phonetic chill or Latin lover effect? Does sound structure or social stereotyping drive FLL? In K. D. Federmeier & H.-W. Huang (Eds.), *Adult and second language learning* (pp. 165–205). Elsevier. https://doi.org/10.1016/bs.plm.2020.02.003

Rendall, D., & Owren, M. J. (2010). Vocalizations as tools for influencing the affect and behavior of others. In S. M. Brudzynski (Ed.), *Handbook of mammalian vocalization: An integrative neuroscience approach* (pp. 177–185). https://doi.org/10.1016/b978-0-12-374593-4.00018-8

Rickford, J. R., & King, S. (2016). Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language*, *92*(4), 948–988. https://doi.org/10.1353/lan.2016.0078

Rosenfelder, M. (2012). *Gen: Language text generator*. Version https://www.zompist.com/gen.html

Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology*, *87*(4), 494–509. https://doi.org/10.1037/0022-3514.87.4.494

Ryan, E. B. (1983). Social psychological mechanisms underlying native speaker evaluations of non-native speech. *Studies in Second Language Acquisition*, *5*(2), 148–159. https://doi.org/10.1017/s0272263100004824

Sebeok, T. A. (2001). *Signs: An introduction to semiotics* (2 ed.). University of Toronto Press.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, *23*(3–4), 193–229. https://doi.org/10.1016/s0271-5309(03)00013-2

Stanley, J. C. (2003). *Tongue of malevolence: A linguistic analysis of constructed fictional languages with emphasis on languages constructed for "the other"*. Duke University. https://doi.org/10.13140/RG.2.2.24534.42561

Stein, S. D. (2023). Space rednecks, robot butlers, and feline foreigners: Language attitudes toward varieties of English in videogames. *Games and Culture*, *18*(8). https://doi.org/10.1177/15554120221150156

Stewart, M. A., Ryan, E. B., & Giles, H. (1985). Accent and social class effects on status and solidarity evaluations. *Personality and Social Psychology Bulletin*, *11*(1), 98–105. https://doi.org/10.1177/0146167285111009

Trudgill, P., & Giles, H. (1976). *Sociolinguistics and linguistic value judgements: Correctness, adequacy and aesthetics*. LAUT.

Vartanian, O., Navarrete, G., Chatterjee, A., Fich, L. B., Leder, H., Modroño, C., Nadal, M., Rostrup, N., & Skov, M. (2013). Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture. *Proceedings of the National Academy of Sciences*, *110*(S2), 10446–10453. https://doi.org/10.1073/pnas.1301227110

Winter, B., Oh, G. E., Hübscher, I., Idemaru, K., Brown, L., Prieto, P., & Grawunder, S. (2021). Rethinking the frequency code: A meta-analytic review of the role of acoustic body size in communicative phenomena. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1840), 20200400. https://doi.org/10.1098/rstb.2020.0400

Winter, B., Pérez-Sobrino, P., & Brown, L. (2019). The sound of soft alcohol: Crossmodal associations between interjections and liquor. *PLoS ONE*, *14*(8). https://doi.org/10.1371/journal.pone.0220449

Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports*, *12*(1), 1035. https://doi.org/10.1038/s41598-021-04311-7

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2.2), 1–27. https://doi.org/10.1037/h0025848