Keep on smiling...? On the sex-specific connections between smiling duration and perceived speaker attributes in business pitches

Lisa M. Tschinse¹, Ali Asadi², Anna Gutnyk³, Oliver Niebuhr²

¹Dept. of Linguistics/Phonetics, University of Cologne, Germany
²IDK/CIE, University of Southern Denmark, Sønderborg, Denmark
³Institute of Linguistic Science and Technology, Nanjing Normal University, Nanjing, China

ltschins@smail.uni-koeln.de, aa@sdu.dk, gutnyk@163.com, olni@sdu.dk

Abstract

We present the results of a perception experiment that uses audio-only speech stimuli to investigate the effect of smiling on speaker traits related to charm and persuasion. Nineteen listeners participated in the experiment and rated excerpts of business pitches given by 18 speakers, 9 males and 9 females, who represented three different groups: non-smilers, occasional smilers, permanent smilers. We find that occasional smilers do best, regardless of gender, when it comes to the charming effect. Men also hardly lose their charm if they smile permanently. The persuasion effect, by contrast, requires fewer smiles. Women should actually not smile at all in order to sound most persuasive. However, the results also show that women already sound more (occasionally) smiling even when they are not smiling at all and, moreover, findings suggest the existence of a perceptual normalization effect for permanently smiling speakers. The results are discussed with regard to their implications for practical presentation coaching and follow-up studies.

Index Terms: smile, prosody, sex, charm, persuasion.

1. Introduction

Smiling is one of the most common expressions of the human body. “From a physical point of view, a smile is a facial expression produced by flexing 17 muscles located around the mouth and the eyes” [1]. But, the smile is more than a mere physical or anatomic action. It is a non-verbal communication signal that consists of both visual (mimic) and acoustic components. We can hear a smile even without seeing it [2,3,4,5,6,7], because the spreading of the lips and the retraction of the corners of the mouth reduce the overall length of the vocal tract, which in turn causes the formant-frequency levels to rise, the formants to move further apart [2,3,8], and the intensity level to increase due to a larger mouth opening [9], see [2,3,4,9]. Also, f0 increases [2,3,8,10] although, unlike for the formant and intensity changes, there is no physical or biomechanical cause for this increase [2] assumes a link between facial and laryngeal tensioning/raising, which is, however, controversial [10].

The link between smiling and an increase in f0 is not deterministic. F0 often does not increases on backchannels although the speaker smiles. That f0 still rises almost automatically even without a physical or biomechanical cause – and thus became a key feature in perceiving smiles [5] – could rather be a residue of the original function of the smile: speakers smile to reduce their acoustical body size, i.e. to appear smaller, more harmless, and more peaceful, also to members of other species [5,11].

Charles Darwin was probably the first to investigate this primeval, cross-species function of the prosodic changes that involve the smile. He described these changes as a “reversed size bluff” [12], in contrast to the opposite changes related to aggression that he called “size bluff”. In the course of human evolution, other functions and forms of smiling were added. An aesthetical dimension emerged as well whose variables are defined and discussed, also with respect to a clear pronunciation, by [1]. Today the form-function relationships of the smile are almost so complex that they represent a language within the language. For example, perceivers can distinguish between up to 16 different types of smiling [13] in visual stimuli, and the set of acoustically distinctive smiles is also large [2,4,5].

In professional life, smiling is often seen as an expression of success and strength. Numerous books on charisma and public speaking allege that smiling has a positive effect, e.g., when giving presentations. For instance, Fox Cabane [14:44] writes “there is only one thing you need to do in order to project more warmth in your voice: smile. [...] it’s worth smiling even when on the phone”. Similarly, Carnegie and Esenwein [15:110] stress in their book The Art of Public Speaking that “the voice with the smile wins”, inspired by the popular slogan of a US telephone company; and Sprague et al. [16] instruct the readers of their Speaker’s Handbook: “Remind yourself to smile genuinely whenever it can reinforce your message” (p. 342), “most speakers underuse [...] this powerful tool” (p.499).

Advice like this implies a positive correlation between smiling and the charismatic, persuasive effect of a speaker. Experimental evidence showing that (genuine) smiles “induce trust and signal higher earning opportunities” [17] is at first glance consistent with such a positive correlation. However, three things speak against the simple, general recommendation that “the voice with the smile [always] wins” [15].

First, we know that the production and perception of speaker charisma are to a certain extent sex-specific. There is a (culture-dependent) gender gap, according to which women have to convey stronger prosodic charismatic signals than men in order to be perceived as equally charismatic [18,19,31]. In addition, there are qualitative differences in what a more charismatic prosody means for men and women. For example, while men have to raise their f0 level to sound more charismatic, women have to lower their f0 level [20]. In relation to smiles, [5,6,21] all emphasize the likelihood of sex differences both in the production and in the perception of male and female smiles. Second, Rosenberg and Hirschberg [22] already assumed that correlations between prosodic parameter levels and perceived speaker charisma are not linear. Niebuhr and colleagues (e.g., [23]) have experimentally confirmed this assumption for practically all prosodic characteristics involved in the perception of charismatic, persuasive speech. So, given that “the more
the better" seem to apply to no other acoustic charisma signal, why should it then apply to the acoustics of the smile?

Third, in connection with the “the more the better” problem, another powerful perceptual mechanism could also come into play, known as perceptual talker normalization, cf. [24,25]. That is, listeners use the previously perceived spectral patterns of a speaker to create a frame of reference that determines how future spectral patterns of that speaker are to be mapped onto phonemes/meanings. For speakers who constantly smile while speaking, the smiled speech could become that spectral frame of reference. In short, this would mean that smiling can best be perceived in contrast to not smiling (like all perceptible properties in hearing and vision rely on contrasts [26]); and it could mean that speakers who permanently smile and lack such a contrast are at a certain point in time no longer perceived as smiling, but only as shorter (due to the increases in f0 and the dispersion and levels of formants [2,3,5,8,11]). Shorter speakers are perceived as less charismatic [27,28, 29]; in this way, permanent smiling could – at least in the auditory domain – reduce rather than support speaker charisma.

Against the outlined background, the aim of our study is to investigate – with audio-only stimuli – the effect of smiling on the perception of male and females speakers. The focus is on the speakers’ charismatic, persuasive effect; but, with reference to the research on aesthetics, we also take a look at the charming effect. Three stimulus sets with different proportions of smiled speech from an equal number of male and female speakers were used to test the (implicit) “the more the better” recommendations for smiling that dominate the advice literature.

We use natural “field” stimuli. The insights that we get from these results are meant to provide the basis for hypotheses that will be tested in a subsequent step with more precisely varied and carefully elicited or ts-generated “lab” stimuli.

2. Method

2.1. Speech material and independent variables

The speech material was extracted from promotional videos uploaded by English teachers to the community-driven language-learning platform italki.com. Anyone who would like to support others in learning a new language in exchange for an hourly fee can upload a video there. The content of these videos is similar to that of a business pitch. The teachers provide some information about themselves and then describe what qualifies them, educationally and professionally, for teaching the respective language. In addition, they outline what special services they can offer to the learners (e.g., flexible bookable hours across time zones, everyday and/or multimedia learning material, or particularly entertaining interactions). The videos typically end with a call-to-action in which the teachers ask the learners to book a course or a free trial lesson with them.

The competition on the italki.com platform is fierce. For EFL (English as a foreign language) alone, more than 8,300 teachers are currently competing for learners. The nature of the videos is correspondingly persuasion-oriented, which makes them ideal for our research question. In addition, all videos have a similar duration of about 90-120 seconds (for such long stimuli, durational differences should not bias trait perception [22]).

We used 18 of these videos, 9 by women and 9 by men, as the basis for our perception experiment. All men and women were native speakers of English and estimated to be 30-50 years old. The 18 videos were selected out of all relevant italki.com videos on the basis of a primarily visual analysis by three independent raters. The task of the three raters was to find videos in which the teachers smile a certain percentage of their speaking time.

Three percentage-ranges, i.e. three categories of teachers were looked for: the non-smilers, the occasional smilers, and the permanent smilers. Teachers who smiled in less than 10 % of their speaking time were considered non-smilers. Occasional smilers smiled between 30-50% of their speaking time. Teachers who smiled for more than 90% of their speaking time were categorized as permanent smilers. Figure 1 shows examples of male and female non-smilers and permanent smilers.

Figure 1: Two m/f non-smilers (top) and two m/f permanent smilers (bottom) whose audio tracks were used as stimuli in the perception experiment.

The selected 18 videos were consistently assigned to one of the three % categories by all three raters and, additionally, had a good and interference-free sound quality. The videos included 6 non-smilers (3m, 3w), 6 occasional smilers (3m, 3w) and 6 permanent smilers (3m, 3w). The 3 proportions of smiled speech represented the independent variable %Smile. A further independent variable was the speaker’s Sex, i.e. male or female.

2.2. Stimulus creation

For stimulus creation, the audio track was extracted from each video as an uncompressed WAV file and saved in mono format. During the first 15-20 seconds of the videos, teachers often gave their name and origin (sometimes also their main job). We cut away this personal information. The cut was made at sentence boundaries that coincided with a topic change. Thus, all videos had a plausible beginning in terms of both content and prosody.

The 18 audio files edited in this way served as stimuli in the experiment without further modifications. Insofar, one can say that the perception experiment involved natural (field) stimuli (business pitches) with a high level of authenticity. The duration of the stimuli varied between 70-90 seconds.

2.3. Participants

Nineteen participants took part in the experiment (26% male and 74% female). They were between 19 and 37 years old (m = 28.52; SD=4.77) and non-native, but fluent speakers of English (according to self-assessments). Thus, the participants exactly matched the target audience of the EFL teachers’ videos.

2.4. Rating scales and dependent variables

The 19 participants were asked to rate the speaker with respect to 9 traits. Two of these traits, pleasant and attractive, were to capture the speaker’s charming effect on listeners. The remaining 7 traits were inspired by the MASCharP scale for perceived charisma [30] and, thus, meant to capture the speaker’s persuasive effect on listeners: convincing, determined, credible, confident, professional, inspiring, and charismatic.
The ratings for all 2+7 traits were made on separate 5-point Likert scales from 1 (‘hardly applies to the speaker’) to 5 (‘fully applies to the speaker’). The Likert-scale ratings were supplemented by two further ones, made by shifting a slider on a continuous percentage scale between 0 and 100. The first of these two supplementary ratings asked the participants to estimate the proportion of smiled speech during the speaker’s presentation, with 0 meaning ‘no smile’ and 100 ‘permanent smile’. The second supplementary rating asked the participants to indicate the likelihood that they would choose the speaker as their EFL teacher; 0 meant ‘definitely no’ and 100 means ‘definitely yes’.

While the percentage ratings were separate dependent variables, the Likert ratings on the 2 or 7 scales for charm and persuasion were merged into one dependent variable each. So, the experiment included four dependent variables (The individual Likert scales became a covariate in the statistics, though).

2.5. Procedure

The perception experiment was conducted as an online questionnaire using the LimeSurvey software. The questionnaire was entitled ‘The sound of good teaching’. Participants were informed that they would have to listen to audio samples of 18 different native-English teachers who introduce themselves to new prospective students and advertise their EFL courses. Based on that, the participants were asked to rate on a set of scales how well each teacher performed in this task. The Likert-scale traits were displayed centered above the corresponding scale. The paired labels like ‘hardly/fully applies’ or ‘definitely yes/no’ were shown next to the ends of the respective scales.

Participants listened to the stimuli via headphones in a silent environment, e.g., their living room (they had to confirm the use of headphones by checking a box on the first page of the questionnaire before they could proceed with the first stimulus).

The 18 stimuli were presented to the participants in individually randomized orders. Completing the whole questionnaire took about 25 minutes.

3. Results

The ratings were statistically analyzed in repeated-measures (RM) ANOVAs based on the two independent variables (fixed factors) %Smile and Sex. Significant between-factor interactions were examined with t-test series, using Sidak corrections to adjust the alpha-error levels for multiple testing.

Four such analyses were conducted, one on the charming effect, one on the persuasion effect, one on the likelihood of choosing the teacher, and one on the estimated degree to which the speaker was smiling during the presentation. In the analyses related to charm and persuasion, Scale was included as a covariate, i.e. 2 two scales/traits associated with charm or the 7 scales/traits associated with persuasion, respectively.

We report the results per RM-AN(C)OVA below, focusing only on the main points because of the limitations of a proceedings paper. For the same reason, the results summary in Figure 2 is restricted to the grand means across the 2 or 7 scales associated with charm or persuasion, respectively.

3.1. Charming effect

The analysis of the speakers’ perceived charm concerned the two scales pleasant and attractive. The corresponding RM-ANOVA shows significant main effects of %Smile (F[2,224] = 21.6, p<.001, η²=.16) and Sex (F[1,112] = 14.6, p<.001, η²=.12). The %Smile*Sex interaction was significant, too (F[2,224]=35.5, p<.001, η²=.24). This overall significance pattern arose from the following key results. Firstly, non-smilers were generally rated less charming than both occasional and permanent smilers. However, it was not the permanent but the occasional smilers who achieved the highest average ratings on the two charm-related scales. Permanent smiling led to a slight decrease in the speakers’ overall charming effect on the listeners (Fig. 2a). Secondly, the latter decrease from occasional to permanent smilers was stronger and only significant for the female speakers. There was no significant decrease for the male speakers. That is, female speakers lost some of their charm when smiling permanently, while male speakers did not (Fig. 2a). Thirdly, male speakers, were rated overall less charming than female speakers (Fig. 2a).

Additionally, we found significant interactions of the covariate Scale (pleasant vs. attractive) with %Smile (F[2,224]= 3.14, p<.001, η²=.22) as well as with Sex (F[1,112] = 65.8, p<.001, η²=.37). The three-way interaction approached significance. The interaction effects of Scale reflect that the advantage in charm of the female speakers over the male speakers was significantly more strongly pronounced for the attractiveness than for the pleasantness scale. Also rating differences as a function of %Smile were generally more pronounced on the attractiveness than on the pleasantness scale. Pleasantness ratings were overall higher than attractiveness ratings, though, especially for non-occasional smilers and male speakers.

3.2. Persuasion effect

The persuasion effect was constituted by the ratings along 7 Likert scales, see above. Like in 3.1, the RM-ANOVA on the persuasion effect yielded significant main effects of %Smile (F[2,794] = 156.4, p<.001, η²=.28) and Sex (F[1,397] = 45.2, p<.001, η²=.11) as well as a significant interaction of %Smile with Sex (F[2,794]=86.7, p<.001, η²=.18). The underlying results pattern is as follows: Firstly, there was, across both sexes, an increase in overall persuasion from permanent smilers to occasional smilers to non-smilers. That is, permanent smiling severely undermined a speaker’s persuasion effect on listeners (Fig. 2b), while fewer smiles supported it. Secondly, male speakers were overall rated more persuasive than female speakers. Thirdly, there was a clear sex-specific smile “tolerance” on part of the listeners. While female speakers were most persuasive when not smiling at all, male speakers achieved overall the highest persuasion ratings with occasional smiles (Fig. 2b). In other words, unlike for the charming effect, women seem to be “punished” for smiling in connection with persuasion. Men, by contrast, are “allowed” to smile, just not permanently.

Effects of the covariate Scale were restricted to an interaction with %Smile (F[2,794]= 19.0, p<.001, η²=.05). The interaction reflects that the %Smile differences described above were more pronounced on some scales (determined, credible, professional) than on others (confident, inspiring).

3.3. Likelihood of choosing the teacher

The RM-ANOVA for teacher choice showed significant main effects of %Smile (F[2,112]=66.3, p<.001, η²=.54) and Sex (F[1,56]=43.5, p<.001, η²=.43) and a significant %Smile*Sex interaction (F[2,112]=67.6, p<.001, η²=.54).

Permanently smiling teachers were 10–20 % less likely to be chosen, and men were about 10 % less likely to be chosen than women. However, similar to the persuasion effect, this
advantage of women applied mainly to those who did not smile at all. So, women were again rewarded for not smiling, whereas men were rewarded for showing occasional smiles (Fig. 2c).

3.4. Estimated degree of smiling

The RM-ANOVA on the perceived percentage of smiled speech in a speaker’s presentation also resulted in a significant main effect of %Smile and Sex (F[2, 112]=961.5, p<.001, \(\eta_p^2=.95\); F[1, 56]=254.9, p<.001, \(\eta_p^3=.82\)). Their two-way interaction was also significant (F[2, 112]=26.8, p<.001, \(\eta_p^3=.32\)).

Two key results must be noted. Firstly, there was no simple linear relationship between the three predefined levels of %Smile selected on a visual basis (see 2.1) and the perceived smiling percentage (0-100%) estimated by the 19 listeners on an auditory basis. The perceived smiling percentage did increase significantly (by about 40%) from non-smilers to occasional smilers (Fig. 2d). Then, however, the estimated degree to which the speaker was smiling during the presentation dropped again significantly by about 10% from occasional to permanent smilers. That is, the highest auditory percentage of perceived smiling was achieved by the occasional smilers, not by the permanent smilers. This applied to both sexes (Fig. 2d).

As regards speaker Sex, we found that women were generally perceived to smile more (Fig. 2d). Even the female non-smilers were already perceived to smile for more than 30% of their presentation time. Moreover, the drop in perceived smiling percentage was significantly larger for the female than for the male speakers. In fact, men who smiled occasionally received a similar perceived smiling percentage (56%) as women who smiled permanently (58%).

![Figure 2: Summary of mean values in the %Smile*Sex interaction of each RM-ANOVA presented in 3.1 (a), 3.2 (b), 3.3 (c), and 3.4 (d); N=57 for each bar.](image)

4. Discussion and Conclusions

Our paper began with a syntactically elliptical question: “Keep on smiling…?”. Our experimental findings suggest answering this question with a clear ‘no’ – in contrast to the rhetoric and advice literature, whose recommendations for smiling seem to underlie a “the more the better” concept. If at all, then such recommendations apply to a speaker’s charming effect, i.e. the degree to which s/he is perceived as pleasant and attractive. But here, too, we see that only men are allowed to smile permanently (i.e. > 90% of the speaking time) in order to appear highly charming. Women, on the other hand, lose their charm (especially their attractiveness) when smiling permanently.

On the other hand, in terms of the persuasive effect, a permanent smile was in our data always detrimental. Thus, as for all other prosodic exponents of perceived speaker charisma [22, 23], there also seems to be an overdose threshold for the (acoustic) smile. It is noteworthy, however, that this threshold seems to be close to 0 for women, at least < 10% according to our definition of non-smilers. Male presenters are allowed to smile significantly more, i.e. about 30-50% of their speaking time.

That women are “punished” for smiling (in business pitches), whereas men are “allowed” to smile, just not permanently, could be a further manifestation of the “gender gap” in perceived speaker charisma. Many studies report such a gap to the disadvantage of women [18, 19, 31, 32]. The main effect of Sex that we found (i.e. also after factoring out all interactions) is indeed consistent with this gap: Men were rated more persuasive/charismatic than women across all scales and %Smile conditions. However, that women should smile less than men (during business pitches) could also have a more basic, auditory reason than the gender gap: Our data indicate a gender bias in the perception of smiled speech. That is, women were generally heard as smiling more than men. Thus, listeners could actually prefer a similar level of %Smile for both sexes, only that women have to smile less than men to reach that level. This gender offset in smiling perception could be related to the higher f0 level of women, i.e. the parameter that differs most strongly between men and women and, at the same time, represents the probably most important acoustic cue to the perception of smiled speech. Future studies need to investigate this question more closely.

Another finding that is worth a second look concerns the drop in ratings from occasional to permanent smilers, which includes the estimated degree of smiling. Permanent smilers do not sound more smiling than occasional smilers. On the contrary, they sound even less smiling. We wonder whether this drop is due to a talker normalization effect [24, 25], caused by an insufficiently strong syntagmatic contrast between smiled and non-smiled intervals in the permanent smilers’ speech signals. That is, perhaps the listeners’ ears got used to the permanent smilers’ smiling speech such that they no longer heard the corresponding acoustic patterns as smiling but reinterpreted them in terms of a smaller (and hence less charismatic) speaker. We will explore this further with controlled stimuli.

Finally, recall that we based our initial study on smiled speech on “field” stimuli. Thus, it is possible (though not likely in view of the replicated previous findings) that effects reported here are merely artifacts of uncontrolled variables in the stimuli. In addition, all reported effects and conclusions refer to audio-only stimuli. It cannot be said to date whether and to what extent A/V stimuli would produce similar results. Nevertheless, at least for negotiations, pitches, or sales talks on the phone, we assume that our results represent valuable practical tips.

5. Acknowledgments

The following link takes you to a conflict-of-interest statement by the last author (ON): https://www.allgoodspeakers.com/coi
6. References


