INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Pitch convergence as an effect of perceived attractiveness and likability

Jan Michalsky[1], *Heike Schoormann*[1]

[1]Institute of German Studies, University of Oldenburg, Germany
j.michalsky@uol.de, heike.schoormann@uol.de

Abstract

While there is a growing body of research on which and how pitch features are perceived as attractive or likable, there are few studies investigating how the impression of a speaker as attractive or likable affects the speech behavior of his/her interlocutor. Recent studies have shown that perceived attractiveness and likability may not only have an effect on a speaker's pitch features in isolation but also on the prosodic entrainment. It has been shown that how speakers synchronize their pitch features relatively to their interlocutor is affected by such impressions. This study investigates pitch convergence, examining whether speakers become more similar over the course of a conversation depending on perceived attractiveness and/or likability. The expected pitch convergence is thereby investigated on two levels, over the entire conversation (globally) as well as turn-wise (locally). The results from a speed dating experiment with 98 mixed-sex dialogues of heterosexual singles show that speakers become more similar globally and locally over time both in register and range. Furthermore, the degree of pitch convergence is greatly affected by perceived attractiveness and likability with effects differing between attractiveness and likability as well as between the global and the local level.

Index Terms: attractiveness, likability, prosodic entrainment, social prosody, convergence, pitch

1. Introduction

Even without a visual impression, a speaker can be perceived as more or less attractive based on acoustic features of the human voice. Amongst those features, pitch constitutes the most prominent and most investigated. Studies found that male speakers were perceived as more attractive when showing lower pitch [1], [2], [3], [4], [5], whereas studies of female voices found that either higher pitch [6], [7], [4], [5] or lower pitch [8], [9] was associated with greater attractiveness. This could either be due to methodological differences or, as suggested by [10], hint at two different concepts of attractiveness for female speakers such as low-pitched sexiness versus high-pitched femininity. Especially in social interaction the overall impression of attractiveness is most likely not only based on the purely visual attractiveness but also affected by the vocal attractiveness.

Additionally, attractiveness may be affected by a person's general likability which is determined by a speaker's behavior as well as influenced by vocal features. However, acoustic features that determine vocal likability are hard to determine since likability itself constitutes a concept that is even harder to grasp than the concept of attractiveness. After all, likability may not constitute a consistent concept but rather serve as a cover term for several personal traits qualifying a person to be perceived as likable [11]. Accordingly, findings are inconsistent and inconclusive. For example, lower pitch has been found to convey warmth [12], pleasantness [13], [14], or likability in general [15], [16] for both male and female speakers, while another study found higher pitch to sound friendlier but only for female speakers [17]. Furthermore, likability has been associated with greater pitch ranges [12], [13] as well as shallower pitch ranges [15].

However, both attractiveness and likability play a role in social interaction. Thus, these two impressions are not only perceived by the interlocutor but may in return also affect his/her vocal behavior to some degree. Accordingly, not only can we expect prosodic effects of attractiveness and likability but also prosodic effects of the perceived impression of attractiveness and likability. While prosodic effects of attraction have been studied and found with male speakers lowering their pitch when talking to more attractive female recipients [18] and female speakers both raising [19] and lowering [18] their pitch compatible with the findings for attractive voices, immediate effects of a speaker's likability on pitch features are scarce. However, in a recent study on perceived attractiveness and likability [20], [21] found both male and female speakers to raise their mean pitch when talking to a more likable interlocutor. In addition, both female and male speakers were found to increase their pitch range when talking to a more attractive interlocutor with no significant effects on the mean pitch and no differences between the sexes.

Besides immediate changes of a speaker's pitch, a phenomenon that receives growing attention in reflecting interpersonal relationships is the so-called *prosodic entrainment* (cf. [15], [22], [23]). Prosodic entrainment describes a speaker adjusting his/her prosodic features with respect to the prosodic features of the respective interlocutor. Speakers can entrain their features either globally across the whole conversation or locally at transition relevant places, where the turn is taken over. Levitan [23], (cf. [24]) distinguishes three types of entrainment. (1) Similarity, where speakers become more similar in absolute terms, (2) synchrony, where speakers adjust their features relatively, thus imitating dynamic changes of the interlocutor without necessarily becoming more similar in absolute terms, and (3) convergence, which describes two speakers becoming more similar over the course of an interaction.

Entrainment was found to occur in more interactively oriented scenarios and is generally regarded as a phenomenon conveying collaboration or rapport [25]. In a previous study, [20], [21] found pitch synchrony in speed dating dialogues both on a global level across conversations as well as on a local level within conversations. Furthermore, the degree of pitch synchrony was significantly affected by both perceived attractiveness as well as likability.

In the study at hand, we investigate pitch convergence in a speed dating setting to answer the question whether the effects

http://dx.doi.org/10.21437/Interspeech.2017-1520

found for pitch synchrony are constant throughout the whole conversation or whether speakers in addition increase their similarity during the course of the conversation. We investigate convergence both globally, comparing the first third and the last third of each conversation, as well as locally on a turn-wise level. Furthermore, we investigate whether global and/or local convergence are affected or even caused by the degree of perceived attractiveness and/or likability. We seek to answer the following research questions:

1) Do speakers decrease the difference between their pitch features and their interlocutor's pitch features over the course of a conversation in a speed dating setting globally and/or locally?

2) Does the perception of attractiveness/likability affect the presence or degree of global or local convergence of a speaker's pitch features?

2. Method

2.1. Speakers

Ten female and 10 male students from the University of Oldenburg participated in the study as paid volunteers. All subjects were monolingual speakers of High German aged between 19 and 28 years. They all grew up in (northern) Germany, i.e. share a common cultural background. Only heterosexual singles were included in the study. All subjects were unacquainted and had no interactions prior to the experiment.

2.2. Procedure

The subjects participated in a speed dating setting, which was altered to meet the research objective. To this end, each participant was paired with each of the 10 participants of the opposite sex resulting in a total of 100 opposite-sex combinations from which 98 could be included in the acoustic analysis. The subjects were placed in a quiet room and instructed to freely engage in a conversation with no restrictions to the topics. A note with sample topics was placed on the table in case participants had difficulties starting the conversation. Each conversation lasted between 15 and 20 minutes. Prior to the first verbal interaction as well as immediately after each conversation, participants received a questionnaire and were asked to evaluate their interlocutor in terms of purely visual attractiveness as well as perceived likability on a 10-point Likert scale. The participants were given the necessary privacy and their ratings were not revealed to the respective interlocutor. Recordings were made in stereo using a portable digital recorder (Tascam HD P2) at a sampling rate of 48 kHz and 24-bit resolution with one head-mounted microphone (DPA 4065 FR) per speaker.

2.3. Acoustic analysis

The acoustic analysis was carried out using Praat [26]. For the analysis of global convergence, audio tracks were separated for each speaker. Subsequently, all filled pauses, laughter, overlaps as well as the interlocutors' speech parts were manually silenced to preserve the time structure of the recordings. All conversations were separated into three equal parts of about five minutes each. The first and the last third of every conversation were used for the acoustic analysis. For local convergence the speech parts of all speakers were segmented into interpausal units [cf. 23]. The two interpausal units adjacent to a turn break with a speaker transition were used for the acoustic analysis. While all 98 conversations were included in the analysis of global effects, the analysis of local effects is based on a subset of 38 conversations. This subset consists of all conversations of four female speakers and about four conversations of each of the ten male speakers. F0 features were extracted from the f0 track of the respective speech parts from both speakers and all measurements calculated in semitones. Subsequently, the differences between the f0 features of both speakers were calculated for the variables shown in table 1.

Table 1: *Acoustic measurements of f0 features.*

Feature	Description
mean	overall f0 mean
median	overall f0 median
max	95th percentile (to exclude outliers)
range1	difference between upper quartile and lower quartile
range2	difference between f0 max and f0 mean

2.4. Statistical analysis

We conducted linear mixed effects models using R [27], the *lme4*-package [28] as well as the *lmerTest*-package [29]. Model fit was determined by maximum likelihood ratio tests. *P*-values were calculated using the Satterthwaite approximation. As fixed effects we used POSITION with the levels *start* and *end* for global convergence and TIME as a continuous variable for local convergence. Furthermore, we included two-way interactions with perceived attractiveness (ATTRACTIVENESS_POST) and likability (LIKABILITY_POST), using the ratings obtained immediately after the respective conversation, as well as the three-way-interaction. As random effects we included random intercepts for *speaker*. The dependent variables were the differences between the speakers regarding the five measurements listed in table 1.

3. Results

3.1. Global convergence

Table 2: *Effects for global convergence on register features.*

	Mean	Median	Max
POSITION	<.001	<.001	<.01
POSITION X ATTRACTIVENESS	<.001	<.001	<.01
POSITION X LIKABILITY	<.05	n.s.	n.s.
POSITION X ATTRACTIVENESS X LIKABILITY	<.01	<.01	<.05

Table 3: *Effects for global convergence on range features.*

	Range1	Range2
POSITION	<.001	n.s.
POSITION X ATTRACTIVENESS	n.s.	n.s.
POSITION X LIKABILITY	<.001	n.s.
POSITION X ATTRACTIVENESS X LIKABILITY	n.s.	n.s.

Table 2 and 3 summarize the significant effects for global convergence. General effects for a global decrease of the differences between two speakers over the course of a

conversation have been found for *mean* (*b*=-2.13, *SE*=0.58, *df*=382.00, *t*=-3.66, ***p*<.001**), *median* (*b*=-2.07, *SE*=0.62, *df*=380.90, *t*=-3.33, ***p*<.001**), *max* (*b*=-1.80, *SE*=0.66, *df*=381.40, *t*=-2.75, ***p*<.01**), *range1* (*b*=-0.34, *SE*=0.08, *df*=389.00, *t*=-4.07, ***p*<.001**) with *range2* being the only exception. Perceived attractiveness significantly increases the degree of global convergence for the register features *mean* (*b*=0.39, *SE*=0.12, *df*=382.10, *t*=3.37, ***p*<.001**), *median* (*b*=0.43, *SE*=0.12, *df*=381.00, *t*=3.50, ***p*<.001**), and *max* (*b*=0.37, *SE*=0.13, *df*=381.50, *t*=2.87, ***p*<.01**) but not for the pitch range features. Perceived likability reaches significance only for increased convergence of *mean* (*b*=0.21, *SE*=0.09, *df*=378.70, *t*=2.42, ***p*<.05**) among the register features. However, perceived likability significantly increases global convergence for *range1* (*b*=0.03, *SE*=0.01, *df*=391.00, *t*=3.55, ***p*<.001**) but not *range2*. Furthermore, interactions between attractiveness and likability have a negative effect on global convergence for *mean* (*b*=-0.04, *SE*=0.02, *df*=379.70, *t*=-2.89, ***p*<.01**), *median* (*b*=-0.04, *SE*=0.02, *df*=378.80, *t*=-2.69, ***p*<.01**), and *max* (*b*=-0.04, *SE*=0.02, *df*=379.20, *t*=-2.23, ***p*<.05**), and no effect for the two range parameters.

3.2. Local convergence

Table 4: *Effects for local convergence on register features.*

	Mean	**Median**	**Max**
TIME	<.01	<.001	<.01
TIME X ATTRACTIVENESS	<.01	<.001	<.001
TIME X LIKABILITY	n.s.	n.s.	<.05
TIME X ATTRACTIVENESS X LIKABILITY	n.s.	n.s.	n.s.

Table 5: *Effects for local convergence on range features.*

	Range1	**Range2**
TIME	n.s.	<.05
TIME X ATTRACTIVENESS	n.s.	<.001
TIME X LIKABILITY	<.05	<.05
TIME X ATTRACTIVENESS X LIKABILITY	<.05	n.s.

Table 4 and 5 summarize the significant effects for local convergence. Local convergence as a main effect is compatible with the effects found for global convergence for *mean* (*b*=-2.56e-04, *SE*=9.55e-05, *df*=790.70, *t*=-2.68, ***p*<.01**), *median* (*b*=-4.18e-04, *SE*=9.24e-05, *df*=791.00, *t*=-4.52, ***p*<.001**), and *max* (*b*=-3.18e-04, *SE*=1.21e-04, *df*=791.20, *t*=-2.63, ***p*<.01**). However, convergence effects on the local level also deviate from global convergence in as far as there are no local effects for *range1* but for *range2* (*b*=-1.23e-04, *SE*=4.92e-05, *df*=501.40, *t*=-2.50, ***p*<.05**). Perceived attractiveness positively affects local convergence for *mean* (*b*=3.96e-05, *SE*=1.47e-05, *df*=789.30, *t*=2.70, ***p*<.01**), *median* (*b*=6.24e-05, *SE*=9.24e-05, *df*=790.10, *t*=4.41, ***p*<.001**), and *max* (*b*=1.01e-04, *SE*=2.13e-05, *df*=791.20, *t*=4.76, ***p*<.001**) similar to global convergence. However, perceived attractiveness also positively affects *range2* (*b*=2.95e-05, *SE*=8.66e-06, *df*=503.30, *t*=3.40, ***p*<.001**). Perceived likability shows unexpected effects as it affects local convergence for *max* (*b*=-4.62e-05, *SE*=2.06e-05, *df*=791.10, *t*=-2.24, ***p*<.05**), *range2* (*b*=-1.83e-05, *SE*=8.47e-06, *df*=710.60, *t*=-2.16, ***p*<.05**), and again for *range1* (*b*=-8.92e-05, *SE*=3.57e-05, *df*=645.90, *t*=-2.50, ***p*<.05**). *Range1*, however, did not show any significant effects for local convergence as a main effect and in all three cases perceived likability negatively affects the degree of entrainment. The only three-way-interaction occurs for *range1* (*b*=1.37e-05, *SE*=6.86e-06, *df*=248.90, *t*=1.99, ***p*<.05**), counteracting the effects of likability on its own.

4. Discussion

First of all, as expected from recent literature on prosodic entrainment [15], [22], [23], the results suggest that speakers show both global and local convergence for most of the investigated pitch features regardless of perceived attractiveness and likability when engaging in an interactive communicative situation such as speed dating. Furthermore, both global and local convergence of most pitch features are greatly affected by both perceived attractiveness and likability.

In contrast to pitch synchrony [20], [21], pitch convergence shows greater effects for perceived attractiveness than for likability. Furthermore, the effects seem nearly complimentary to the effects found for pitch synchrony. On the global level, perceived attractiveness affects the entrainment of all three register parameters but shows no effects on the pitch range. Perceived likability, however, shows significant effects for one of the register parameters (the mean) only but affects the entrainment of the pitch range. In the previous findings on global synchrony, the effect of perceived attractiveness was limited to the pitch range and the effect of perceived likability was limited to the register [20], [21].

On the local level the differences become even clearer. The effects of perceived attractiveness are consistent and significantly enhance local convergence of all three register features. Furthermore, attractiveness also enhances the convergence of the range from maximum to mean which was neither significant as a main effect nor in interaction with attractiveness on the global level. Perceived likability however does not affect the convergence of mean nor median f0 thus deviating from the global effects. However, likability affects the convergence of the max and both range features, but in the opposite direction by decreasing the degree of convergence for all three parameters.

Conclusively, the study shows that speakers converge over the course of a conversation both globally across greater stretches of speech as well as locally at adjacent turns. The finding that both global and local convergence are dependent on the speaker's impression of their interlocutor regarding visual attractiveness and overall likability holds several implications.

First of all, it suggests that prosodic entrainment seems to be more than a mere automatic categorical adaptation to an interlocutor in social interaction but is instead dependent on social variables (cf. [30], [31], [32], [23]). This adds to the previous findings of pitch synchrony which was also found to be dependent on the perceived attractiveness and likability both globally and locally [21].

Secondly, taking the effects for pitch synchrony into account [20], [21], speakers both show an increase in overall similarity over time and dynamically adjust their pitch features to those of their interlocutors. Both forms of entrainment are affected by perceived attractiveness and likability, sometimes resulting in contradicting effects for the exact same pitch features [20], [21]. This draws a very complex picture of the pitch effects in social interaction and requires further investigation.

Thirdly, as for pitch synchrony, the effects for pitch convergence differ greatly between the global and the local level thus supporting the assumption that the first interpausal unit after a speaker transition serves a prominent function in signaling social features via pitch [21].

Lastly, by investigating entrainment for each speaker and not the conversation as a whole, and by furthermore finding that entrainment is affected by the individual perception of attractiveness and likability we can assume that prosodic entrainment does at least to a certain degree not constitute a balanced mutual process but both speakers can entrain to different degrees. This even allows for the possibility of prosodic entrainment to be completely one-sided with only one speaker entraining in the direction of the other while the interlocutor shows no entrainment at all.

5. References

[1] S. A. Collins, "Men's voices and women's choices," *Animal Behaviour*, vol. 60, pp. 773–780, 2000.

[2] D. R. Feinberg, L. M. Debruine, B. C. Jones, and D. I. Perrett, "Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices," *Animal Behaviour*, vol. 69, pp. 561–568, 2005.

[3] C. R. Hodges-Simeon, S. J. C. Gaulin, and D. A. Puts, "Different vocal parameters predict perceptions of dominance and attractiveness," *Human Nature*, vol. 21, pp. 406–427, 2010.

[4] B. C. Jones, D. R. Feinberg, L. M. Debruine, A. C. Little, and J. Vukovic, "A domain- specific opposite-sex bias in human preferences for manipulated voice pitch," *Animal Behaviour*, vol. 79, no. 57-62, 2010.

[5] Y. Xu, A. Lee, W.-L. Wu, X. Liu, and P. Birkholz, "Human Vocal Attractiveness as Signaled by Body Size Projection," *PLoS ONE*, vol. 8, no. 4, 2013.

[6] S. A. Collins and C. Missing, "Vocal and visual attractiveness are related in women," *Animal Behaviour*, vol. 65, pp. 997–1004, 2003.

[7] D. R. Feinberg, L. M. Debruine, B. C. Jones, and D. I. Perrett, "The role of femininity and averageness of voice pitch in aesthetic judgements of women's voices," *Perception*, vol. 37, pp. 615–623, 2008.

[8] T. Oguchi and H. Kikuchi, "Voice and interpersonal attraction," *Japanese Psychological Research*, vol. 39, pp. 56–61, 1997.

[9] K. Leaderbrand, J. Dekam, A. Morey, and L. Tuma, "The effects of voice pitch on perceptions of attractiveness: Do you sound hot or not," *Winona State University Psychology Student Journal*, 2008.

[10] A. Karpf. *The human voice*. New York, NY: Bloomsbury Publishing, 2006.

[11] B. Weiss, "Akustische Korrelate von Sympathieurteilen bei Hörern gleichen Geschlechts," 26th Konferenz Elektronische Sprachsignalverarbeitung (ESSV), Eichstädt. Studientexte zur Sprachkommunikation. TUD Press, Dresden, vol. 78, pp. 165–171, 2015.

[12] B. L. Brown, W. J. Strong, and A. C. Rencher, "Fifty-four voices from two: the effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech," *Journal of the Acoustical Society of America*, vol. 55, no. 2, pp. 313–318, 1974.

[13] L. Bruckert, J. Lienard, A. Lacroix, M. Kreutzer, and G. Leboucher, "Women use voice parameter to assess men's characteristics," *Proceedings in Biological Sciences*, vol. 237, pp. 83–89, 2006.

[14] B. Ketzmerick, "Zur auditiven und apparativen Charakterisierung von Stimmen," *Studientexte zur Sprachkommunikation. TUD Press, Dresden*, 2007.

[15] A. Gravano, R. Levitan, L. Willson, S. Beňuš, J. Hirschberg, and A. Nenkova, "Acoustic and prosodic correlates of social behavior," *Proceedings of Interspeech*, 2011.

[16] B. Weiss, "Prosodische Elemente vokaler Sympathie," *Studientexte zur Sprachkommunikation. TUD Press, Dresden*, vol. 65, pp. 212–217, 2013.

[17] D. Jurafsky, R. Ranganath, and D. McFarland, "Extracting social meaning: Identifying interactional style in spoken conversation," *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics*, pp. 638–646, 2009.

[18] S. M. Hughes, S. D. Farley, and B. C. Rhodes, "Vocal and physiological changes in response to the physical attractiveness of conversational partners," *Journal of Nonverbal Behavior*, vol. 34, pp. 1–13, 2010.

[19] P. J. Fraccaro, B. C. Jones, J. Vukovic, F. G. Smith, C. D. Watkins, D. R. Feinberg, A. C. Little, and L. M. Debruine, "Experimental eivdence that women speak in higher voice pitch to men they find attractive," *Journal of Evolutionary Psychology*, vol. 9, no. 1, pp. 57–67, 2011.

[20] J. Michalsky and H. Schoormann, "Effects of perceived attractiveness and likability on global aspects of fundamental frequency," *Proceedings of P&P12*, pp. 120–124, 2016.

[21] J. Michalsky, "Pitch synchrony as an effect of perceived attractiveness and likability," *Proceedings of DAGA 2017*, 2017.

[22] R. Levitan, A. Gravano, L. Willson, S. Beňuš, J. Hirschberg, and A. Nenkova, "Acoustic-prosodic entrainment and social behavior," *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 11–19, 2012.

[23] R. Levitan. Acoustic-prosodic entrainment in human-human and human-computer dialogue. Columbia University. PhD thesis, 2014.

[24] J. Edlund, M. Heldner, and J. Hirschberg, "Pause and gap length in face-to-face interaction," *Proceedings of Interspeech*, 2009.

[25] N. Lubold and H. Pon-Barry, "Acoustic-Prosodic Entrainment and Rapport in Collaborative Learning Dialogues," Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge, November 12-12, 2014, Istanbul, Turkey, 2014.

[26] P. Boersma and D. Weenink. *Praat: Doing phonetics by computer*, 2016.

[27] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[28] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[29] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen. lmerTest: Tests in linear mixed effects models. R package version 2.0-30, 2016.

[30] M. Natale, "Convergence of mean vocal instenity in dyadic communication as a function of social desirability," *Journal of Personality and Social Psychology*, vol. 32, no. 5, pp. 790–804, 1975.

[31] H. Giles, N. Coupland, and J. Coupland, "Accomodation theory: Communication, context, and consequence. Contexts of accomodation," *Developments in applied sociolinguistics*, vol. 1, 1991.

[32] T. L. Chartrand and J. A. Bargh, "The chameleon effect: The perception-behavior link and social interaction," *Journal of Personality and Social Psychology*, vol. 76, no. 6, pp. 893–910, 1999.