



OPEN ACCESS

EDITED BY

Maurizio Codispoti,
University of Bologna, Italy

REVIEWED BY

Jianqiang Zhang,
Jiangnan University, China
Mingyu Joo,
University of California, Riverside,
United States

*CORRESPONDENCE

Mizuki Nagano
✉ mizuki.nagano@ntt.com

SPECIALTY SECTION

This article was submitted to
Emotion Science,
a section of the journal
Frontiers in Psychology

RECEIVED 09 August 2022

ACCEPTED 01 December 2022

PUBLISHED 09 January 2023

CITATION

Nagano M, Ijima Y and Hiroya S (2023)
Perceived emotional states mediate
willingness to buy from advertising
speech. *Front. Psychol.* 13:1014921.
doi: 10.3389/fpsyg.2022.1014921

COPYRIGHT

© 2023 Nagano, Ijima and Hiroya. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction
in other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Perceived emotional states mediate willingness to buy from advertising speech

Mizuki Nagano^{1*}, Yusuke Ijima¹ and Sadao Hiroya²

¹NTT Human Informatics Laboratories, NTT Corporation, Yokosuka, Japan, ²NTT Communication Science Laboratories, NTT Corporation, Atsugi, Japan

Previous studies have shown that stimulus-organism-response (SOR) theory can well explain the willingness to buy from stores, products, and advertising-related stimuli. However, few studies have investigated advertising speech stimulus that is not influenced by visual design. We examined whether SOR theory using emotional states can explain the willingness to buy from advertising speech stimulus. Participants listened to speech with modified speech features (mean F0, speech rate, and standard deviation of F0) and rated their willingness to buy the advertised products and their perceived emotional states (pleasure, arousal, dominance). We found that emotional states partially mediate the influence of speech features on the willingness to buy. We further analyzed the moderating effects of listeners' attributes and found that listeners' gender and age group moderated the relationship between speech features, emotional states, and willingness to buy. These results indicate that perceived emotional states mediate the willingness to buy from advertising speech.

KEYWORDS

willingness to buy, advertising speech, emotional states, SOR theory, PAD model, age difference, gender difference

1. Introduction

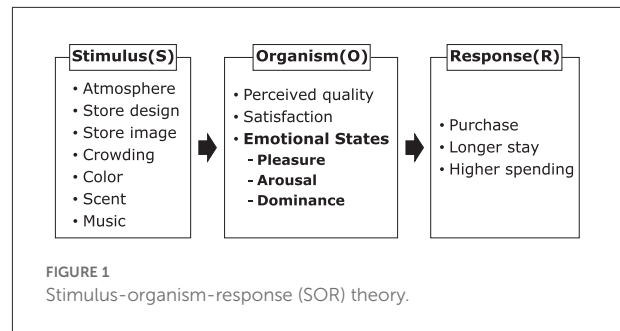
Retailers and manufacturers always make an effort to attract as many customers as possible and increase sales. They pay attention to many factors regarding their stores and products, e.g., store environment, layout, product packaging, and advertising. These factors make customers interested in the product and increase their willingness to buy.

It is well known that atmosphere, crowding, background music, image, glossiness, haptic features, and speech influence the willingness to buy (Kotler, 1973; Turley and Milliman, 2000; Krishna, 2012). In consumer behavior research, stimulus-organism-response (SOR) theory as a hierarchical model was proposed to explain the willingness to buy from these stimuli and is better than directly explaining the willingness to buy from the stimuli (Mehrabian and Russell, 1974). It is suggested that high-level processing in the brain may be involved in determining the willingness to buy from stimuli.

In SOR theory, a model that uses pleasure, arousal, and dominance (PAD) emotional states as an organism is called the PAD model (Donovan and Rossiter, 1982). This model has successfully been used to explain the effects of store-related stimuli, such as crowding and background music, and product-related stimuli, such as glossiness and haptic features, on the willingness to buy (Mari and Poggesi, 2013; Briand Decré and Cloonan, 2019). There has also been research on the effect of advertising-related stimuli, such as copy, image, video, and speech, on the willingness to buy. For example, previous studies reported on the copy and image of catalogs and video advertisements (Fiore, 2002; Wu and Chen, 2016). There is a consistently positive correlation between the pleasure emotional state and willingness to buy (Donovan et al., 1994; Anwar et al., 2020).

Speech is also used as advertising-related stimuli, especially for effective broadcasting advertising. However, there is not a very consistent relationship between speech features and the willingness to buy. For example, Chattopadhyay et al. have showed that a faster speech rate and low pitches have influenced the willingness to buy (Chattopadhyay et al., 2003), while Peterson et al. have showed that there was no correlation between mean F0 and SD F0 of salespeople and performance (Peterson et al., 1995). Since speech is closely related to emotional states and known to affect the perceived impression or behavior of the listener (Schröder, 2001; Tsantani et al., 2016; Li and Akagi, 2018), such a nonlinear relationship between speech features and the willingness to buy may be explained using a hierarchical model. In fact, in their study on the willingness to buy, Poon et al. did not use the PAD model but investigated the influence of the differences in pause length of male and female speech on the willingness to buy through a perceived personality state (Poon et al., 2018). However, there have been few studies on the PAD model and taking into account speech as a stimulus, including the willingness to buy as a response.

We examined whether the PAD model can explain the willingness to buy from the speech stimulus of electric appliance advertisements spoken by male and female professional narrators. To verify the effect of the difference in the speech features, the speech was generated with the converted static and dynamic features of F0 and speech rate, which have been used to study speech features and emotional states (Schröder, 2001). These speech features can be manipulated with simple speech signal processing. Large-scale subjective evaluations were conducted *via* crowdsourcing. Participants were asked to evaluate their emotional states, i.e., PAD and their willingness to buy when listening to an advertising speech. We conducted an analysis of variance (ANOVA), path analysis, mediation analyzes, and moderated mediation analysis using the obtained data of the speech features of speech stimuli, emotional states, and the willingness to buy.

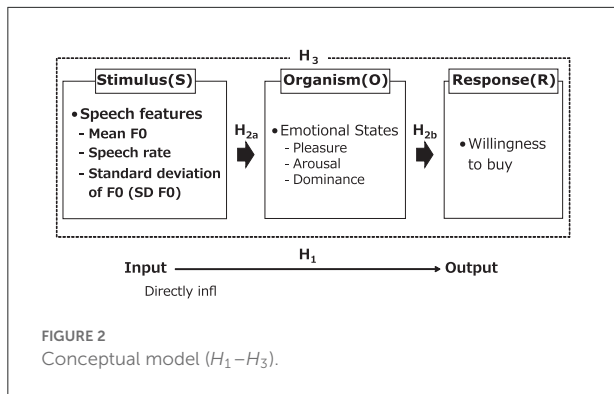


2. Stimulus-organism-response (SOR) theory and hypotheses

SOR theory (Mehrabian and Russell, 1974) comprises three dimensions: stimulus (S), organism (O), and response (R) (Figure 1). Stimulus incorporates all external environmental factors in the store such as atmosphere (Donovan et al., 1994), design (Jang et al., 2018; Nusairat et al., 2020), brand image (Simanjuntak et al., 2020), crowding (Anninou et al., 2018), color, scent, and music (Roschk et al., 2017). These stimuli stimulate various consumer responses. In previous studies, the responses were investigated using perceived quality (Nusairat et al., 2020), satisfaction (Roschk et al., 2017), and emotional states (Anwar et al., 2020). These consumer responses result in approach or avoidance behavior. Approach behavior refers to a positive attitude toward the environment, such as staying in a place. In contrast, avoidance behavior refers to a negative attitude toward the environment, such as escaping from a place. In consumer behavior research, approach or avoidance behavior is confirmed with indicators such as purchase, longer stay time (Jang et al., 2018) and higher spending (Milliman, 1982).

The PAD model specifically focuses on emotion as the organism in SOR theory. There are three emotional states: PAD (Donovan and Rossiter, 1982). Pleasure refers to the degree of feeling joy, satisfaction, and happiness with a situation. Arousal refers to the degree of feeling excited, passionate, and active about the situation. Dominance refers to the degree of feeling that an individual has influence over a situation and can control it. The PAD model has been verified with various external stimuli, and the evidence from previous studies has shown that this model reflects the relations among these stimuli, consumers' emotional states, and consumer' responses (Donovan and Rossiter, 1982; Milliman, 1982; Donovan et al., 1994; Mari and Poggesi, 2013; Briand Decré and Cloonan, 2019; Anwar et al., 2020). We expect the PAD model to also explain the willingness to buy caused by speech stimuli.

Thus, the paper attempts to verify the conceptual model (Figure 2) and the following four hypotheses:



H_1 : Speech features directly influence the willingness to buy (S-R). No specific hypothesis about speech features is given because of the lack of consistent results in previous studies.

H_{2a} : Speech features influence emotions (S-O).

H_{2b} : Emotional states influence the willingness to buy (O-R). In particular, when pleasure increases, the willingness to buy increases.

H_3 : Emotional states mediate the influence of speech features on the willingness to buy (SOR).

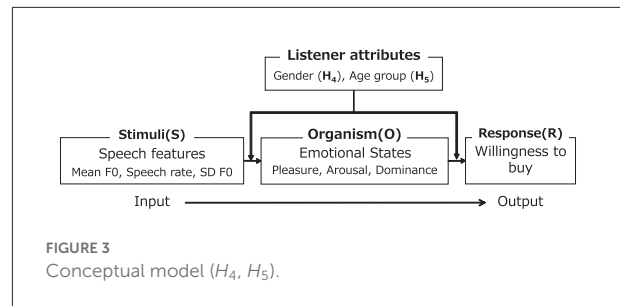
The main purpose of this study was to verify H_3 , and the verification of H_1 and H_2 is to confirm the preconditions of H_3 .

However, the effects of emotional mediating on the willingness to buy from advertising speech may vary depending on attributes such as age and gender of the listener. For example, Schirmer et al. (2005) reported that female participants were more capable of recognizing emotions than male participants. Coley and Burgess (2003) report that the effect of emotional states on impulse buying tends to be more evident for females than with males. Regarding age differences, it is well known that older people have a tendency to preferentially process positive stimuli relative to negative stimuli, called the positive effect (Reed and Carstensen, 2012), and that hearing loss with age (Dupuis and Pichora-Fuller, 2015). Older people have also been reported to have lower emotion recognition abilities than younger adults (Lima et al., 2014; Schmidt et al., 2016). Therefore, the relationship between advertising speech, emotional states, and the willingness to buy may be moderated by the attributes of the listener. At least, it is expected that the attributes of the listener will influence the relationship between emotion and the willingness to buy, which is consistent with previous studies.

We further attempted to verify the conceptual model (Figure 3) and the following two hypotheses:

H_4 : The relationship between advertising speech and emotional states or between emotion and the willingness to buy is moderated by the gender of the listener.

H_5 : The relationship between advertising speech and emotional states or between emotion and the willingness to buy is moderated by the age group of the listener.



3. Subjective evaluation

3.1. Speech material

Four sentences of read speech stimuli spoken by one Japanese male and one Japanese female professional narrators were used. All sentences were gathered from the Web and intended to promote electrical appliances that customers cannot recognize much of a difference between brands for products (Assael, 1984). Since all of the advertising sentences we used were about older products, we also concluded that listeners are less likely to be influenced in their willingness to buy by the advantage of the products advertised in the sentences. Advertising speech for four electrical appliances was used: air conditioner, washing machine, refrigerator, and PC. Sentences translated into English are included in the Supplementary material.

Both narrators spoke the same four sentences. The average F0, standard deviation of F0 (SD F0), sentence duration, and speech rate were 112.87 [Hz], 1.34 [Hz], 15.41 [s], and 6.57 [mora/s] for the male speech and were 255.53 [Hz], 1.30 [Hz], 14.87 [s], and 7.02 [mora/s] for the female speech, respectively. The sampling frequency of the recorded speech was 22.05 [kHz]. Each phoneme boundary was manually segmented. We attempted to minimize the difference in SD F0, pause duration, and speech rate between male and female speech by recording speech, but in addition to the mean F0, there was a significant difference in the speech rate between genders [mean F0: $t_{(6)} = 25.61, p < 0.05$, SD F0: $t_{(6)} = 0.89, p = 0.41$, speech rate: $t_{(6)} = 5.16, p < 0.05$].

To verify the influence of speech features on emotional states and the willingness to buy, we manipulated the mean F0 [Hz], SD F0, and speech rate [mora/s] of the original speech. Mean F0 was converted by a factor of 0.94 (low) or 1.06 (high) to the average F0 of the sentence using WORLD (Morise et al., 2016), and speech rate was converted by a factor of 1.12 (slow) or 0.89 (fast) by PICOLA (pointer interval controlled overlap and add) (Morita and Itakura, 1986). The conversion rate for mean F0 was determined on the logarithmic axis to take into account human auditory characteristics. SD F0 was converted by a factor of 1.50 (more variation) or 0.67 (less variation) (Fukuoka et al., 2016). These parameters were determined through

TABLE 1 Participants' age groups and gender.

| | 20s | 30s | 40s | 50s | 60s | Total |
|--------|-----|-----|-----|-----|-----|-------|
| Male | 48 | 49 | 45 | 52 | 53 | 247 |
| Female | 35 | 43 | 33 | 20 | 22 | 153 |
| Total | 83 | 92 | 78 | 72 | 75 | 400 |

preliminary experiments so that the converted speech would not sound unnatural compared with human speech. To reduce the influence of sound quality deterioration due to speech modification mentioned above, all stimuli used in the subjective evaluation were analyzed and synthesized using WORLD: a process of analysis-by-synthesis by WORLD was also conducted even for the original speech used as a baseline. We used 27 types of speech stimuli per sentence in the combination of mean F0, speech rate, and SD F0 [3 (mean F0: low vs. original vs. high) \times 3 (speech rate: slow vs. original vs. fast) \times 3 (SD F0: large vs. original vs. small)]. Pause duration was matched by adjusting the silent sections between male and female speech. The average intensity level of all stimuli was 62 dB.

3.2. Participants

The participants were 457 native Japanese speakers. They were recruited *via* a crowdsourcing service and participated in the experiment on their judgment. Fifty-seven participants had to be excluded from the analysis due to incomplete answers. The remaining 400 participants were available for analysis (247 males and 153 females, mean age = 42.48 years, $SD = 14.29$, range = 21–70). The participants' age group and gender are shown in Table 1. All participants were paid after their completion of the experiment.

3.3. Procedure

Each participant listened to the speech of either a male or female narrator. Two of the four sentences (54 stimuli) were selected for each participant, taking counterbalance into account. The experiment was conducted in a browser *via* crowdsourcing. The participants listened to the speech stimuli only once, rated the degree of their willingness to buy the advertised product, then rated their perceived emotional states in the browser. The order of evaluated items was designed on the basis of a previous study (Milliman, 1982). The willingness-to-buy response was removed from the display when the participants answered regarding their emotional states. Because of the crowdsourcing experiment, it was difficult to control the playback equipment and presented sound-pressure level among the participants, so they participated at a comfortable volume for them. Note that there is little difference between the

results of auditory laboratory experiments and crowdsourcing experiments (Cooke and García Lecumberri, 2021).

We aimed to investigate the effects of changing the speech features of the same speaker's speech on emotional states and willingness to buy. Therefore, the speakers of the advertising speech were limited, but we considered all combinations of changes in speech features.

The willingness to buy was rated on a 7-point Likert scale (1: not at all willing to buy–7: very willing to buy). Participants were given the following instructions: "You don't really have to think about whether or not to buy it because we want you to evaluate "motivation." Please answer not how you feel about the manufacturer or brand, but how you feel about the narrator's way of speaking." The participant's emotional states were rated on a 7-point Likert scale: [pleasure (pleasant–unpleasant), arousal (calm–excited), and dominance (dominant–submissive)]. Participants were given the instructions used in a previous study (Mori et al., 2005) to make it easier to understand these emotional state dimensions (e.g., "Pleasure refers to how good or bad you feel"). Participants answered their age group and gender at the end of the experiment. The experiment lasted about 30 min.

3.4. Analysis

For all analyzes, the evaluation values for emotional state and the willingness to buy collected in the experiment were used. The mean F0 [Hz], SD F0 [Hz], and speech rate [mora/s] in the speech interval of each stimulus were used as speech features. Since there were significant differences between the male and female speech in mean F0 and speech rate, which were obtained by subtracting the average values of each were used in the analysis below.

To analyze the direct influence of speech features on the willingness to buy (H_1), a three-way ANOVA was conducted. The significance of the main effects of speech features (mean F0, speech speed, and SD F0) and their interactions were verified. We should confirm that speech features need to influence the willingness to buy before conducting the following path and mediation analyzes.

Path analysis was conducted to examine direct dependencies between speech features and emotions (H_{2a}) and between emotions and willingness to buy (H_{2b}). Path analysis is a commonly used analysis method in consumer-behavior

research (Nusairat et al., 2020). It can verify hypothesized models by analyzing direct and indirect relationships between multiple variables. The goodness-of-fit index indicates how well the hypothesized model fits the experimental data. The degree of the effect between variables can be compared using the standardized path coefficient.

Path analysis is insufficient for verifying the importance of considering emotional states to estimate the willingness to buy from speech features. Thus, mediation analysis was used to verify whether emotional states mediate the influence of speech features on the willingness to buy (H_3). With this analysis, it is assumed that a mediator M has a potential influence between an independent variable X and dependent variable Y when X affects Y (Rijnhart et al., 2021). There should be a causal relationship between X and Y as well as between X and M and, M and Y . This analysis is used to verify how M contributes. The effect of X on Y is called the total effect t . The effect of X when adjusted for M is called the direct effect d . The mediating effect e is defined as $t - d$. Each statistical significance was concluded with 1,000 bootstrap samples. The effect is considered significant if it does not include zero within the 95% confidence interval calculated from the bootstrap samples. When t is significant and d is not significant, M is regarded to be completely mediating the effect of X on Y and is called “complete mediation.” When both t and d are significant, M is regarded to be partly mediating the effect of X on Y and is called “partial mediation.”

The speech features were entered as independent variables, emotional states as the mediating variable, and the willingness to buy as the dependent variable. In Figure 4, these effects are shown as the following equations.

$$e = t - d = P_a \times P_b + A_a \times A_b + D_a \times D_b$$

Because variables must have causal relationships, only significant paths were used in the analysis on the basis of the results of the path analysis.

We conducted a moderated mediation analysis to examine whether the gender and age group of the listener moderated the pathways shown in Figure 3 in the mediation process of speech features on willingness to buy through emotional states (H_4 , H_5 ; Edwards and Lambert, 2007; Hayes, 2015). We calculated the moderated e of speech features on the willingness to buy and the conditional indirect effect at specific levels of the gender or age group when emotional states as a mediator. To avoid the problem of multicollinearity, each variable was centered. In the conditional indirect effect, the representative values of the moderators (e.g., mean, mean±SD, or dummy variable) were selected (Aiken et al., 1991; Kim and Bae, 2019). Each statistical significance was concluded with 1,000 bootstrap samples as with mediation analysis. The effect is considered significant if it does not include zero within the 95% confidence interval calculated from the bootstrap samples.

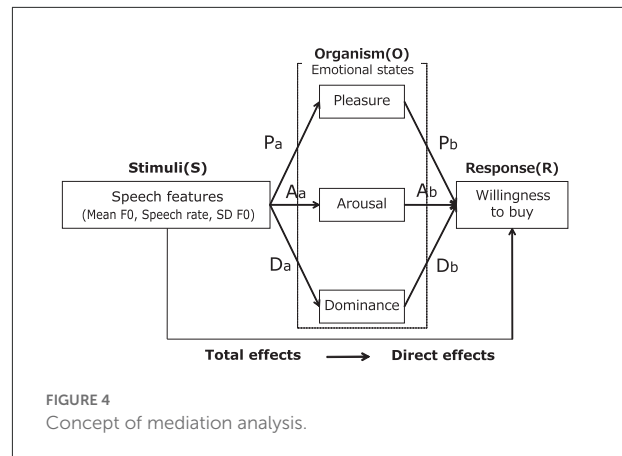


FIGURE 4
Concept of mediation analysis.

4. Results of applicability of PAD model using advertising speech

4.1. Three-way ANOVA

A three-way ANOVA was conducted to analyze the direct influence of speech features on the willingness to buy. Table 2 shows the mean and variance for the evaluation values used in the analysis. There were main effects of mean F0 [$F_{(2, 21,573)} = 39.66, p < 0.01$], speech rate [$F_{(2, 21,573)} = 1870.59, p < 0.01$], and SD F0 [$F_{(2, 21,573)} = 529.60, p < 0.01$]. The first-order interaction effects were observed between speech rate and SD F0 [$F_{(4, 21,573)} = 4.55, p < 0.01$], but not between mean F0 and speech rate [$F_{(4, 21,573)} = 2.17, p = 0.07$] or between mean F0 and SD F0 [$F_{(4, 21,573)} = 1.99, p = 0.09$]. The significant first-order interaction effect was analyzed for simple main effects. The results indicate that each factor was significant. The faster speech rate enhanced the willingness to buy under all SD F0 conditions [under less condition: $F_{(2, 7,197)} = 676.30, p < 0.01$; under original condition: $F_{(2, 7,197)} = 686.10, p < 0.01$; under more condition: $F_{(2, 7,197)} = 515.00, p < 0.01$]. More SD F0 enhanced the willingness to buy under all speech rate conditions [under slow condition: $F_{(2, 7,197)} = 130.60, p < 0.01$; under original condition: $F_{(2, 7,197)} = 225.00, p < 0.01$; under fast condition: $F_{(2, 7,197)} = 129.40, p < 0.01$]. The second-order interaction was not observed [$F_{(8, 21,573)} = 0.93, p = 0.49$]. The result revealed that mean F0, speech rate, and SD F0 influence the willingness to buy. Thus, H_1 was supported. High mean F0, fast speech rate, or large SD F0 tended to increase the willingness to buy.

4.2. Path analysis

Path analysis (Duncan, 1966) was conducted to examine direct dependencies between speech features and emotional states and between emotional states and the willingness to

TABLE 2 Evaluation values used in analysis.

| Variable | Mean | SD |
|--------------------|------|------|
| Pleasure | 4.01 | 1.29 |
| Arousal | 3.76 | 1.29 |
| Dominance | 3.94 | 1.06 |
| Willingness to buy | 3.88 | 1.24 |

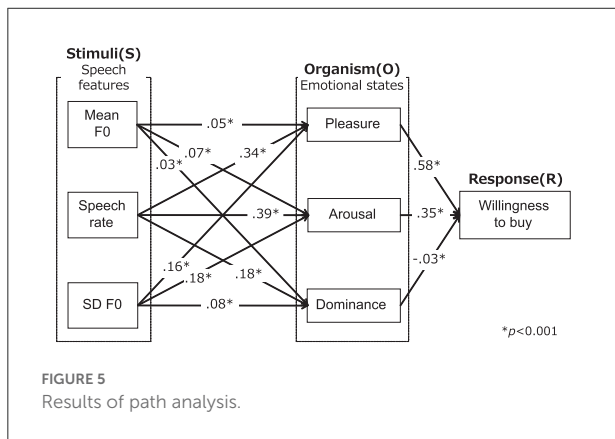


FIGURE 5 Results of path analysis.

TABLE 3 Goodness-of-fit indices for path analysis.

| Fit indices | Accepted value | Value |
|-----------------------|----------------|--------|
| χ^2 (Chi-square) | | 121.13 |
| df | | 6 |
| GFI | >0.9 | 0.999 |
| AGFI | >0.9 | 0.997 |
| CFI | <0.10 | 0.997 |
| RMSEA | <0.10 | 0.003 |

GFI, AGFI, CFI, and RMSEA, respectively, indicate goodness of fit index, adjusted GFI, comparative fit index, and root mean square error of approximation.

buy. Figure 5 shows a path diagram with a path coefficient. Some fit indices in Table 3 indicate that the model is valid. As shown in Figure 5, all speech features had a significant positive effect on all dimensions of the emotional states. Thus, H_{2a} was supported. The speech rate had a greater path coefficient than other speech features. For the effect of the emotional states on the willingness to buy, all emotional-state dimensions were significant. Therefore, H_{2b} was supported.

4.3. Mediation analysis

Figure 6 and Table 4 show the results of mediation analysis. The mediating effects of PAD on the willingness to buy were all significant because the 95% confidence interval did not include zero. The largest mediating effect was pleasure, followed in order

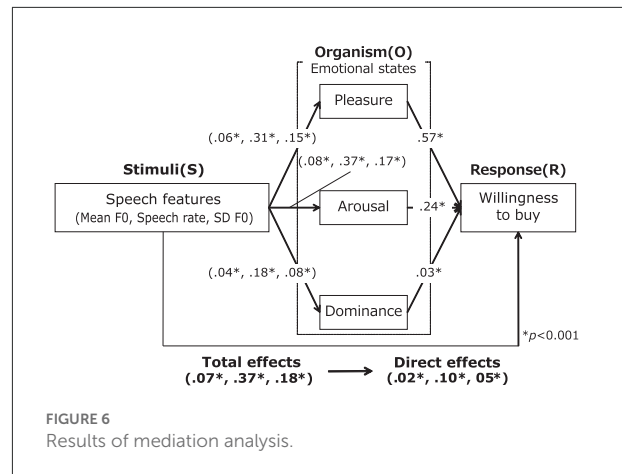


FIGURE 6 Results of mediation analysis.

TABLE 4 Mediating effect of emotional states.

| | Pleasure | Arousal | Dominance |
|-----------|----------|---------|-----------|
| Estimates | 1.52 | 0.75 | 0.04 |
| SE | 0.05 | 0.03 | 0.01 |
| CI.Lower | 1.42 | 0.69 | 0.02 |
| CI.Upper | 1.63 | 0.82 | 0.06 |

by arousal and dominance. The total and direct effects of speech features on the willingness to buy were both significant. These results indicate that PAD have a partial mediating effect on the relationship between speech features and the willingness to buy and that the emotion-mediated model is applicable to advertising speech. This supports hypothesis H_3 .

5. Moderating effects of listener's gender and age group

The results of the moderated mediation analysis by listeners' gender are shown in Table 5. In an outcome of emotional states, all speech features had significant positive effects on all emotional states, and the product of speech rate and gender and that of SD F0 and gender for pleasure and arousal had significant positive effects. In an outcome of the willingness to buy, all speech features and emotional states had significant positive effects on the willingness to buy. The product of pleasure and gender and that of dominance and gender had significant positive effects, and that of arousal and gender had significant negative effects.

Table 6 shows the results of the conditional mediating effect of emotional states when gender moderated between speech features and emotion, and between emotional states and the willingness to buy. All mediating effects of emotional states were significant for female listeners. For male listeners, the mediating effects of pleasure and arousal were significant, but

TABLE 5 Results of moderated mediation analysis for listeners' gender.

| Outcome | Speech features→Emotional states | | | | | | | | | Emotional states →Willingness to buy | | |
|--------------------|----------------------------------|------|-------|---------|------|-------|-----------|------|-------|---|------|-------|
| | Pleasure | | | Arousal | | | Dominance | | | B | SE | t |
| | B | SE | t | B | SE | t | B | SE | t | B | SE | t |
| Mean F0 | 0.32** | 0.03 | 9.49 | 0.40** | 0.03 | 12.03 | 0.17** | 0.03 | 5.65 | 0.08** | 0.02 | 3.69 |
| Speech rate | 0.62** | 0.01 | 49.0 | 0.73** | 0.01 | 59.77 | 0.30** | 0.01 | 27.37 | 0.19** | 0.01 | 21.65 |
| SD F0 | 1.86** | 0.08 | 23.79 | 2.09** | 0.08 | 27.48 | 0.82** | 0.07 | 12.13 | 0.59** | 0.05 | 12.31 |
| Gender | 0.18** | 0.02 | 10.67 | 0.07** | 0.02 | 4.06 | -0.04** | 0.01 | -2.75 | -0.37** | 0.04 | -8.26 |
| Mean F0×Gender | 0.05 | 0.07 | 0.70 | 0.12 | 0.07 | 1.71 | 0.04 | 0.06 | 0.71 | | | |
| Speech rate×Gender | 0.07** | 0.03 | 2.84 | 0.16** | 0.03 | 6.49 | 0.04 | 0.02 | 1.58 | | | |
| SD F0×Gender | 0.32* | 0.16 | 1.98 | 0.70** | 0.16 | 4.45 | 0.21 | 0.14 | 1.54 | | | |
| Pleasure | | | | | | | | | | 0.55** | 0.01 | 72.73 |
| Arousal | | | | | | | | | | 0.24** | 0.01 | 30.15 |
| Dominance | | | | | | | | | | 0.03** | 0.01 | 4.49 |
| Pleasure×Gender | | | | | | | | | | 0.05** | 0.02 | 3.33 |
| Arousal×Gender | | | | | | | | | | -0.04** | 0.02 | -2.36 |
| Dominance×Gender | | | | | | | | | | 0.06** | 0.01 | 4.26 |
| F | 448.9 | | | 642.5 | | | 133.1 | | | 4409 | | |
| R ² | 0.13** | | | 0.17** | | | 0.04** | | | 0.67** | | |

Columns 2–4 and 5 show moderating effects of gender on emotional states and on willingness to buy, respectively. * $p < 0.05$ and ** $p < 0.01$.

the mediating effect of dominance was not significant because the 95% confidence interval included zero. The mediating effect of all emotional states tended to be larger for female listeners than for male listeners. Thus, H_4 was supported.

The results of the moderated mediation analysis by listeners' age group are shown in Table 7. In an outcome of emotions, all speech features had significant positive effects on all emotional states. The product of speech rate and age group for all emotional states and that of SD F0 and age group in arousal and dominance had significant positive effects. In an outcome of willingness to buy, all speech features and emotional states had significant positive effects on the willingness to buy. The product of pleasure and age group had significant positive effects, and that of arousal and age group and that of dominance and age group had significant negative effects on the willingness to buy.

Table 8 shows the results of the conditional mediating effect of emotional states when the age group moderated between speech features and emotional states, and between emotional states and the willingness to buy. This comparison was conducted at three levels of the listener's age group: mean -1SD (younger), mean, and mean +1SD (older). The mediating effect of pleasure was significant in all conditions and tended to be larger as listeners got older. The mediating effect of arousal was significant under all conditions and tended to be smaller as listener age increased, unlike pleasure. The mediating effect of dominance was significant for young and mean age but did not differ. For older, the mediating effect of dominance was not significant because the 95% confidence interval included zero. This suggests that speech that enhances pleasure in older people and arousal in younger people may increase their willingness to buy. Thus, H_5 was supported.

6. Discussion

This study was mainly motivated to investigate whether SOR theory mediated by emotional states can explain the willingness to buy from advertising speech. From the results of ANOVA, we found that speech features of mean F0, speech rate, and SD F0 affected the willingness to buy (H_1). Path analysis confirmed that there was a dependency between speech features and emotional states and between emotional states and the willingness to buy (H_{2a} and H_{2b}). Through mediation analysis, we found that there was a mediating effect of emotional states between speech features and the willingness to buy (H_3). These results suggest that SOR theory using the PAD model can explain the willingness to buy from advertising speech.

The result that fast-speech rate increases the willingness to buy is consistent with Chattopadhyay (Chattopadhyay et al., 2003)'s results. However, our results were not consistent with their finding of increased willingness to buy at a low mean F0 (Chattopadhyay et al., 2003) or with the finding of no correlation between mean F0 and SD F0 of salespeople and

TABLE 6 Conditional mediating effect of emotional states when gender moderated between speech features and emotional states, and between emotional states and willingness to buy.

| | Pleasure | | | Arousal | | | Dominance | | | | | |
|---------------|-----------|-------|-------|---------|-------|-------|-----------|-------|-------|------|-------|------|
| | Estimates | 95%CI | | SE | 95%CI | | Estimates | 95%CI | | | | |
| | | Lower | Upper | | Lower | Upper | | Lower | Upper | | | |
| Male | 1.40 | 0.06 | 1.28 | 1.52 | 0.72 | 0.04 | 0.64 | 0.79 | 0.01 | 0.01 | -0.01 | 0.03 |
| Female | 1.77 | 0.08 | 1.61 | 1.94 | 0.80 | 0.06 | 0.69 | 0.92 | 0.09 | 0.02 | 0.06 | 0.12 |
| Not-moderated | 1.52 | 0.05 | 1.42 | 1.63 | 0.75 | 0.03 | 0.69 | 0.82 | 0.04 | 0.01 | 0.02 | 0.06 |

TABLE 7 Results of moderated mediation analysis for listeners' age.

| Outcome | Speech features→Emotional states | | | | | | | | | Emotional states →Willingness to buy | | |
|-----------------------|----------------------------------|------|-------|---------|------|-------|-----------|------|-------|---|------|--------|
| | Pleasure | | | Arousal | | | Dominance | | | B | SE | t |
| | B | SE | t | B | SE | t | B | SE | t | | | |
| Mean F0 | 0.32** | 0.03 | 9.48 | 0.40** | 0.03 | 12.01 | 0.17** | 0.03 | 5.66 | 0.07** | 0.02 | 3.65 |
| Speech rate | 0.62** | 0.01 | 48.88 | 0.73** | 0.01 | 59.69 | 0.30** | 0.01 | 27.41 | 0.18** | 0.01 | 22.79 |
| SD F0 | 1.85** | 0.08 | 23.54 | 2.08** | 0.08 | 27.34 | 0.82** | 0.07 | 12.20 | 0.60** | 0.05 | 12.80 |
| Age group | -0.01 | 0.01 | -1.24 | 0.02** | 0.01 | 2.79 | 0.00 | 0.01 | -0.36 | 0.00 | 0.00 | -1.30 |
| Mean F0×Age group | 0.00 | 0.02 | -0.09 | -0.03 | 0.02 | -1.31 | -0.04 | 0.02 | -1.77 | | | |
| Speech rate×Age group | 0.04** | 0.01 | 4.70 | 0.03** | 0.01 | 3.25 | 0.05** | 0.01 | 6.14 | | | |
| SD F0×Age group | 0.05 | 0.06 | 0.87 | 0.11* | 0.05 | 2.02 | 0.26** | 0.05 | 5.39 | | | |
| Pleasure | | | | | | | | | | 0.56** | 0.01 | 104.38 |
| Arousal | | | | | | | | | | 0.22** | 0.01 | 38.56 |
| Dominance | | | | | | | | | | 0.03** | 0.01 | 5.34 |
| Pleasure×Age group | | | | | | | | | | 0.06** | 0.01 | 15.15 |
| Arousal×Age group | | | | | | | | | | -0.03** | 0.01 | -6.22 |
| Dominance×Age group | | | | | | | | | | -0.01** | 0.01 | -3.18 |
| F | 432.4 | | | 632.9 | | | 141.7 | | | 4454 | | |
| R ² | 0.12** | | | 0.17** | | | 0.04** | | | 0.67** | | |

Columns 2–4 and 5 show moderating effects of age on emotional states and on willingness to buy, respectively. * $p < 0.05$ and ** $p < 0.01$.

TABLE 8 Conditional mediating effect of emotional states when age moderated between speech features and emotional states and willingness to buy.

| | Pleasure | | | | Arousal | | | | Dominance | | | |
|---------------|-----------|------|-------|-------|-----------|------|-------|-------|-----------|------|-------|-------|
| | Estimates | SE | 95%CI | | Estimates | SE | 95%CI | | Estimates | SE | 95%CI | |
| | | | Lower | Upper | | | Lower | Upper | | | Lower | Upper |
| Mean - 1SD | 1.31 | 0.05 | 1.20 | 1.41 | 0.78 | 0.05 | 0.69 | 0.88 | 0.04 | 0.01 | 0.02 | 0.06 |
| Mean | 1.56 | 0.05 | 1.45 | 1.66 | 0.72 | 0.03 | 0.66 | 0.78 | 0.04 | 0.01 | 0.02 | 0.06 |
| Mean + 1SD | 1.82 | 0.06 | 1.70 | 1.95 | 0.64 | 0.04 | 0.57 | 0.73 | 0.02 | 0.02 | -0.01 | 0.06 |
| Not-moderated | 1.52 | 0.05 | 1.42 | 1.63 | 0.75 | 0.03 | 0.69 | 0.82 | 0.04 | 0.01 | 0.02 | 0.06 |

Mean - 1SD, Mean, and Mean + 1SD are 25, 39, and 53 years old, respectively.

sales performance by Peterson et al. (1995). One reason for this difference could be that their emotional states were different. Another possible reason is the difference in experimental methods between speech conversion in this study and speech analysis of multiple salespeople in Peterson et al.'s study.

Rosenberg et al. examined the relationship between the rating of some speakers' charisma for political statements and several speech features (Rosenberg and Hirschberg, 2009). They reported that a higher mean F0, greater SD F0, and faster speech rate were rated more charismatic, which is the same trend as in our study. Although the rating of charisma is a different task from the willingness to buy, the reason the results indicate the same trend may be due to the unconscious response of wanting to buy products from trustworthy persons.

As a result of the path analysis, the path coefficients from speech rate to emotional states and from pleasure to the willingness to buy were the largest. These results are consistent with previous studies (Tursunov et al., 2019; Anwar et al., 2020; Nagano et al., 2021).

The results of the mediation analysis showed that emotional states had a partial mediating effect because the significance of the direct effect of the speech features was observed. If the mediating effect of the emotional states was sufficient as a mediator, the significance of the direct effect should be lost. Thus, there are other mediators besides emotional states regarding the effect of speech features on the willingness to buy. Another possible mediating factors is impressions of speech. It may be effective to subdivide the relationship between speech features and emotional states (Li and Akagi, 2018).

The emotional mediating effects on the willingness to buy from advertising speech may differ depending on the attributes of the listeners. We analyzed the moderating effects of the listener's gender and age group on emotional states and the willingness to buy.

In moderated mediation analysis, the mediating effects of all emotional states differed depending on the listeners' gender (H4). It tended to be larger for female listeners than for male listeners. This may be related to the fact that females have better emotional perception than males (Bonebright et al., 1996; Kring and Gordon, 1998; Keshtiari and Kuhlmann, 2016).

From the results for the listeners' age group, the mediating effects of all emotional states, in particular, pleasure and arousal, differed depending on the listeners' age group (H5). The mediating effect of pleasure tended to be larger as listeners age increased, but the mediating effect of arousal tended to be smaller. Schmidt et al. compared the emotional recognition ability for various speakers' emotional voices between younger and older adults (Füllgrabe et al., 2015). They reported that older adults were more likely than younger adults to respond to differences in mean F0 cueing pleasure and less strongly to intensity differences cueing arousal. The results of our study are similar. It is also possible that hearing loss in older adults may have affected the results, but further investigation is required.

When the moderating effects of listeners' gender and age group were taken into account, the mediating effect of dominance for male and older were not significant. The significance of dominance is still controversial, with many studies indicating there is no significant difference (Donovan and Rossiter, 1982; Donovan et al., 1994), and others indicating there is (Yalcha and Spangenberg (2000)). The effect of dominance may be more likely affected by individual characteristics such as listeners' attributes.

The finding that emotional states mediate the willingness to buy from advertising speech and that the attributes of the listener moderate the mediating effect suggests that salespeople may improve sales by speaking in a way that appeals to the customer's emotional states. However, our study had several other limitations. Since the experiment was conducted *via* crowdsourcing, the listening environment (e.g., device, intensity, background noise) of the participants was not controlled. These differences in the experiment may have influenced the results.

Another problem is that the types of speech features, sentences, the domain of advertisement, and speakers were limited, and participants were not required to buy the products promoted within the advertising speech. To confirm whether the results of this study are generally applicable, experiments with more speakers, items, or sentences and participants' real purchasing behavior should be conducted (Joo et al., 2020).

7. Conclusion

We analyzed the relationship between speech features, emotional states (pleasure, arousal, dominance), and the willingness to buy on the basis of a consumer-behavior model. Large-scale subjective evaluation data were collected *via* crowdsourcing. The participants listened to speech with different speech features (mean F0, speech rate, or SD F0) and rated their willingness to buy the products advertised in the speech and their perceived emotional states. The results of a three-way ANOVA indicate that speech features affect the willingness to buy. As the result of path and mediation analyzes, the emotional states were revealed to function as a partial mediator regarding the influence of speech features on the willingness to buy, and the emotion-mediated model is effective. In particular, increasing pleasure and arousal can be expected to enhance the willingness to buy. The moderating effects of the listener's gender and age group on emotional states and the willingness to buy were also analyzed. The mediating effects of all emotional states tended to be larger for female listeners than for male listeners. From the results of the listeners' age group, the mediating effect of pleasure tended to be larger as listener age increased, but the mediating effect of arousal tended to be smaller. For future work, we will investigate whether the same tendency is observed when speech sentences or speakers are changed.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval were not required for this study in accordance with the national legislation and institutional requirements. Written informed consent for participation was not required in accordance with national legislation and institutional requirements.

Author contributions

MN and YI conceived the study and carried out the experiment. MN, YI, and SH carried out the modeling and analytical investigations. All authors wrote the paper and reviewed the manuscript and contributed to the article and approved the submitted version.

Acknowledgments

The authors wish to thank all the participants who took part in the experiments.

Conflict of interest

MN, YI, and SH were employed by company NTT Corporation.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1014921/full#supplementary-material>

References

- Aiken, L. S., West, S. G., and Reno, R. R. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: Sage Publications.
- Anninou, I., Stavradi, G., and Yu, Y. (2018). "Cultural differences on perceived crowding, shopping stress and excitement in superstores," in *Proceedings of the 51th Academy of Marketing Conference* (Stirling), 1–13.
- Anwar, A., Waqas, A., Zain, H. M., and Kee, D. M. H. (2020). Impact of music and colour on customers' emotional states: an experimental study of online store. *Asian J. Bus. Res.* 10, 104–125. doi: 10.14707/ajbr.200077
- Assael, H. (1984). *Consumer Behavior and Marketing Action*. Boston, MA: Kent Publication Co.
- Bonebright, T. L., Thompson, J. L., and Leger, D. W. (1996). Gender stereotypes in the expression and perception of vocal affect. *Sex Roles.* 34, 429–445. doi: 10.1007/BF01547811
- Briand Décré, G., and Cloonan, C. (2019). A touch of gloss: haptic perception of packaging and consumers' reactions. *J. Product Brand Manag.* 28, 117–132. doi: 10.1108/JPBMM-05-2017-1472
- Chattopadhyay, A., Dahl, D. W., Ritchie, R. J. B., and Shahin, K. N. (2003). Hearing voices: the impact of announcer speech characteristics on consumer response to broadcast advertising. *J. Consum. Psychol.* 13, 198–204. doi: 10.1207/S15327663JCP1303_02
- Coley, A., and Burgess, B. (2003). Gender differences in cognitive and affective impulse buying. *J. Fashion Market. Manag.* 7, 282–295. doi: 10.1108/13612020310484834
- Cooke, M., and García Lecumberri, M. L. (2021). How reliable are online speech intelligibility studies with known listener cohorts? *J. Acoust. Soc. Am.* 150, 1390–1401. doi: 10.1121/1.5005880
- Donovan, R., and Rossiter, J. (1982). Store atmosphere: an environmental psychology approach. *J. Retail.* 58, 34–57.
- Donovan, R. J., Rossiter, J. R., Marcooly, G., and Nesdale, A. (1994). Store atmosphere and purchasing behavior. *J. Retail.* 70, 283–294. doi: 10.1016/0022-4359(94)90037-X
- Duncan, O. D. (1966). Path analysis: sociological examples. *Am. J. Sociol.* 72, 1–16. doi: 10.1086/224256
- Dupuis, K., and Pichora-Fuller, M. K. (2015). Aging affects identification of vocal emotions in semantically neutral sentences. *J. Speech Lang. Hear. Res.* 58, 1061–1076. doi: 10.1044/2015_JSLHR-H-14-0256
- Edwards, J. R., and Lambert, L. S. (2007). Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. *Psychol. Methods* 12, 1–22. doi: 10.1037/1082-989X.12.1.1
- Fiore, A. M. (2002). Effects of experiential pleasure from a catalogue environment on approach responses toward fashion apparel. *J. Fashion Market. Manag.* 6, 122–133. doi: 10.1108/13612020210429467
- Fukuoka, I., Takatsu, H., Fujie, S., Hayashi, Y., and Kobayashi, T. (2016). "Prosodic analysis for utterance sequence in spoken dialogue based information," in *Proceedings of the 30th Annual Conference of the Japanese Society for Artificial Intelligence* (Fukuoka), 1–4.
- Füllgrabe, C., Moore, B. C. J., and Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front. Aging Neurosci.* 6, 347. doi: 10.3389/fnagi.2014.00347
- Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behav. Res.* 50, 1–22. doi: 10.1080/00273171.2014.962683
- Jang, J. Y., Baek, E., Yoon, S. Y., and Choo, H. J. (2018). Store design: visual complexity and consumer responses. *Int. J. Design* 12, 105–118.
- Joo, M., Liu, W., and Wilbur, K. C. (2020). Divergent temporal courses for liking versus wanting in response to persuasion. *Emotion* 20, 261–270. doi: 10.1037/emo0000543
- Keshtiar, N., and Kuhlmann, M. (2016). The effects of culture and gender on the recognition of emotional speech: evidence from persian speakers living in a collectivist society. *Int. J. Soc. Cult. Lang.* 4, 71–86. doi: 10.13140/RG.2.1.115.9.0001
- Kim, E., and Bae, S. (2019). Gratitude moderates the mediating effect of deliberate rumination on the relationship between intrusive rumination and post-traumatic growth. *Front. Psychol.* 10, 2665. doi: 10.1007/978-981-32-9721-0
- Kotler, P. (1973). Atmospherics as a marketing tool. *J. Retail.* 49, 48–64.
- Kring, A. M., and Gordon, A. H. (1998). Sex differences in emotion: expression, experience, and physiology. *J. Pers. Soc. Psychol.* 74, 686–703. doi: 10.1037/0022-3514.74.3.686
- Krishna, A. (2012). An integrative review of sensory marketing: engaging the senses to affect perception, judgment and behavior. *J. Consum. Psychol.* 22, 332–351. doi: 10.1016/j.jcps.2011.08.003
- Li, X., and Akagi, M. (2018). A three-layer emotion perception model for valence and arousal-based detection from multilingual speech. *Proc. Interspeech* 2, 3643–3647. doi: 10.21437/Interspeech.2018-1820
- Lima, C. F., Alves, T., Scott, S. K., and Castro, S. L. (2014). In the ear of the beholder: how age shapes emotion processing in nonverbal vocalizations. *Emotion* 14, 145–160. doi: 10.1037/a0034287
- Mari, M., and Poggesi, S. (2013). Servicescape cues and customer behavior: a systematic literature review and research agenda. *Service Ind. J.* 33, 171–199. doi: 10.1080/02642069.2011.613934
- Mehrabian, A., and Russell, J. A. (1974). *Approach to Environmental Psychology*. Cambridge, MA: The MIT Press.
- Milliman, R. E. (1982). Using background affect to music behavior of the supermarket shoppers. *J. Market.* 46, 86–91. doi: 10.1177/002224298204600313
- Mori, H., Aizawa, H., and Kasuya, H. (2005). Consistency and agreement of paralinguistic information annotation for conversational speech. *J. Acoust. Soc. Jpn.* 61, 690–697. doi: 10.20697/jasj.61.12_690
- Morise, M., Yokomori, F., and Ozawa, K. (2016). World: a vocoder-based high-quality speech synthesis system for real-time applications. *IEICE Trans. Inf. Syst.* E99-D, 1877–1884. doi: 10.1587/transinf.2015EDP7457
- Morita, N., and Itakura, F. (1986). "Time-scale modification algorithm for speech by use of pointer interval control overlap and add (PICOLA) and its evaluation," in *Proceedings of Annual Meeting of Acoustical Society of Japan* (Akita), 149–150.
- Nagano, M., Ijima, Y., and Hiroya, S. (2021). "Impact of emotional state on estimation of willingness to buy from advertising speech," in *Proceedings of Interspeech* (Brno), 2486–2490.
- Nusairat, N., Hammouri, Q., Al-Ghadir, H., Ahmad, A. M. K., and Eid, M. A. H. (2020). The effect of design of restaurant on customer behavioral intentions. *Manag. Sci. Lett.* 10, 1929–1938. doi: 10.5267/j.msl.2020.2.021
- Peterson, R., Cannito, M., and Brown, S. (1995). An exploratory investigation of voice characteristics and selling effectiveness. *J. Pers. Sell. Sales Manag.* 15, 1–15.
- Poon, M., Chan, K., and Yiu, E. (2018). "The relationship between speech rate, voice quality and listeners' purchase intentions," in *Proceedings of the 9th International Conference on Speech Prosody* (Poznan), 468–472.
- Reed, A. E., and Carstensen, L. L. (2012). The theory behind the age-related positivity effect. *Front. Psychol.* 3, 339. doi: 10.3389/fpsyg.2012.00339
- Rijnhart, J. J., Valente, M. J., MacKinnon, D. P., Twisk, J. W., and Heymans, M. W. (2021). The use of traditional and causal estimators for mediation models with a binary outcome and exposure-mediator interaction. *Struct. Equat. Model.* 28, 345–355. doi: 10.1080/10705511.2020.1811709
- Roschk, H., Loureiro, S. M. C., and Breitsohl, J. (2017). Calibrating 30 years of experimental research: a meta-analysis of the atmospheric effects of music, scent, and color. *J. Retail.* 93, 228–240. doi: 10.1016/j.jretai.2016.10.001
- Rosenberg, A., and Hirschberg, J. (2009). Charisma perception from text and speech. *Speech Commun.* 51, 640–655. doi: 10.1016/j.specom.2008.11.001
- Schirmer, A., Striano, T., and Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport* 16, 635–639. doi: 10.1097/00001756-200504250-00024
- Schmidt, J., Janse, E., and Scharenborg, O. (2016). Perception of emotion in conversational speech by younger and older listeners. *Front. Psychol.* 7, 781. doi: 10.3389/fpsyg.2016.00781
- Schröder, M. (2001). "Emotional speech synthesis: a review," in *Proceedings of the 7th European Conference on Speech* (Aalborg), 561–564.
- Simanjuntak, M., Nur, H. R., Sartono, B., and Sabri, M. F. (2020). A general structural equation model of the emotions and repurchase intention in modern retail. *Manag. Sci. Lett.* 10, 801–814. doi: 10.5267/j.msl.2019.10.017
- Tsantani, M. S., Belin, P., Paterson, H. M., and McAleer, P. (2016). Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices. *Perception* 45, 946–963. doi: 10.1177/0301006616643675

Turley, L. W., and Milliman, R. E. (2000). Atmospheric effects on shopping behavior: a review of the experimental evidence. *J. Bus. Res.* 49, 193–211. doi: 10.1016/S0148-2963(99)00010-7

Tursunov, A., Kwon, S., and Pang, H. S. (2019). Discriminating emotions in the valence dimension from speech using timbre features. *Appl. Sci.* 9, 1–18. doi: 10.3390/app9122470

Wu, Y.-L., and Chen, P.-C. (2016). “The synesthesia effects of online advertising stimulus design on word-of-mouth and purchase intention: from the perspective of consumer olfactory and gustatory,” in *AMCIS* (San Diego, CA).

Yalcha, R. F., and Spangenberg, E. R. (2000). The effects of music in a retail setting on real and perceived shopping times. *J. Bus. Res.* 49, 139–147. doi: 10.1016/S0148-2963(99)00003-X