Can AI Powered Speech-to-Text and Text-to-Speech techniques limit the interviewer bias in sensory and consumer research?

Hester Kreuzen, Daniëlle Dull, Vera de Rover, Rignald Span
EyeQuestion Software (Logic8 bv), Elst, The Netherlands

Introduction & Objectives

Using a human interviewer to ask panellists about their product evaluation and opinion can lead to several types of interviewer bias. These factors lead to a difference between the participant's answer and the true value. On average this difference has the same direction for all participants, generating a systematic error.[1]

Types of interviewer bias that can occur are:

- Social desirability bias → tendency to give socially desirable answers[2]
- Acquiescence bias → tendency to give positive answers[3]
- Effect of a health claim → tendency to give positive evaluations to products with a health claim[4]

A potential way to reduce these biases is to use Speech-to-Text and Text-to-Speech (StT) techniques.

➢ **Is there a difference in bias when a human interviewer is used compared to when StT techniques are used?**

Methods

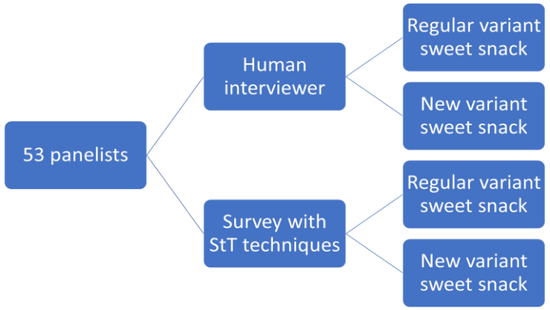

Figure 1. Overview of the study design

58 Dutch consumers were recruited from the Blindgetest panel. 53 panelists completed both test conditions (17 males and 36 females; average age 45 years).

Both conditions took place at the panellists' home. The human interview was done via a videocall using Zoom. In the StT survey the instructions and questions were read out loud by the computer. Panellists answered the questions by using their voice. Questionnaires were created using EyeQuestion Software (version 5.0.4.4, 2021).

In each condition two sweet snacks were evaluated: a regular variant (no health claim) and a new variant (same regular product but with health claim) and panellists were asked for their preference. Food consumption behaviour was also assessed. Conditions and products within each condition were balanced.

A satisfaction score (table 1) and unhealthiness score (table 2) was calculated based on the attributes and questions that were included in the questionnaires.

Attribute	Question(s)	Minimum score	Maximum score
Appearance	Appearance hedonic scale (1 – 9)	1	9
Odor	Odor hedonic scale (1 – 9)	1	9
Taste	Taste hedonic scale (1 – 9) + Sweetness JAR (-2 – 0 – -2)	1 + -2 = -1	9 + 0 = 9
Texture	Texture hedonic scale (1 – 9) + Softness/firmness JAR (-2 – 0 – -2)	1 + -2 = -1	9 + 0 = 9
Aftertaste	Aftertaste hedonic scale (1 – 9) + Aftertaste intensity JAR (-2 – 0 – -2)	1 + -2 = -1	9 + 0 = 9
General opinion	Overall liking hedonic scale (1 – 9) + Purchase intent scale (1 – 5) + Recommendation intent scale (1 – 5)	1 + 1 + 1 = 3	9 + 5 + 5 = 19
Overall opinion	Total score	1 + 1 + -1 + -1 + -1 + 3 = 2	9 + 9 + 9 + 9 + 9 + 19 = 64

Table 1. Corresponding questions, and minimum and maximum scores for each attribute for the satisfaction score.

Unhealthiness score		
Question(s)	**Minimum score**	**Maximum score**
Frequency of sweets consumption (1 – 9) + Frequency of savories consumption (1 – 9) + Frequency of fruits and vegetables consumption (1 – 9) + Importance of health (1 – 5) + Importance of vitamins and minerals (1 – 5) + Importance of low fats and sugars (1 – 5)	1 + 1 + 1 + 1 + 1 + 1 = 6	9 + 9 + 9 + 5 + 5 + 5 = 42

Table 2. Corresponding questions, and minimum and maximum score for the unhealthiness score.

References

1. Lavrakas, P. J. (2008). Response Bias. In *Encyclopedia of Survey Research Methods*. Retrieved from: https://dx.doi.org/10.4135/9781412963947.n486
2. Hall, M. F. (1995). Patient satisfaction or acquiescence? Comparing mail and telephone survey results. *Journal of health care marketing, 15(1),* 54-61.
3. Krumpal, I. (2011). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity, 47(4),* 2025–2047. https://doi.org/10.1007/s11135-011-9640-9
4. Sabbe, S., Verbeke, W., Deliza, R., Matta, V., & Van Damme, P. (2009). Effect of a health claim and personal characteristics on consumer acceptance of fruit juices with different concentrations of açaí (Euterpe oleracea Mart.). *Appetite, 53(1),* 84–92. https://doi.org/10.1016/j.appet.2009.05.014

Results

Acquiescence bias

	Regular product			"New" product		
	Interview	StT	P-value	Interview	StT	P-value
Appearance	7.15	7.11	0.359	**7.40**	**6.98**	**0.012**
Odor	6.47	6.60	0.800	**6.64**	**5.89**	**0.044**
Taste	6.40	6.47	0.772	**7.00**	**6.32**	**0.011**
Texture	6.38	6.36	0.955	6.75	6.49	0.357
Aftertaste	6.04	6.00	0.894	**6.74**	**5.72**	**0.003**
General opinion	13.30	13.58	0.522	14.47	13.02	0.059
Overall opinion	45.74	46.13	0.748	**49.00**	**44.42**	**0.001**

Table 3. Mean satisfaction score for the different attributes in the two conditions, split by product. Values of attributes with a significant p-value (<0.05) are displayed in bold.

Effect of the health claim

	Human interviewer			StT survey		
	Regular	"New"	P-value	Regular	"New"	P-value
Appearance	**7.15**	**7.40**	**0.001**	7.11	6.98	0.322
Odor	6.47	6.64	0.642	6.60	5.89	0.471
Taste	**6.40**	**7.00**	**0.030**	6.47	6.32	0.630
Texture	6.38	6.75	0.159	6.36	6.49	0.623
Aftertaste	**6.04**	**6.74**	**0.024**	6.00	5.72	0.496
General opinion	**13.30**	**14.47**	**0.021**	13.58	13.02	0.842
Overall opinion	**45.74**	**49.00**	**0.021**	46.13	44.42	0.782

Table 4. Mean satisfaction score for the different attributes for the two products, split by condition. Values of attributes with a significant p-value (<0.05) are displayed in bold.

Preference

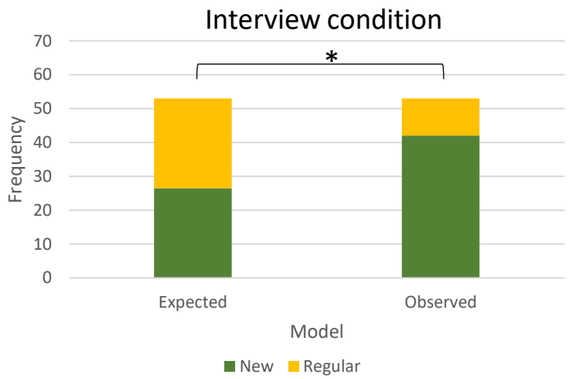

Figure 2. Frequencies of choosing the "new" or regular product for the expected model based on chance and the observed model in the interview condition.

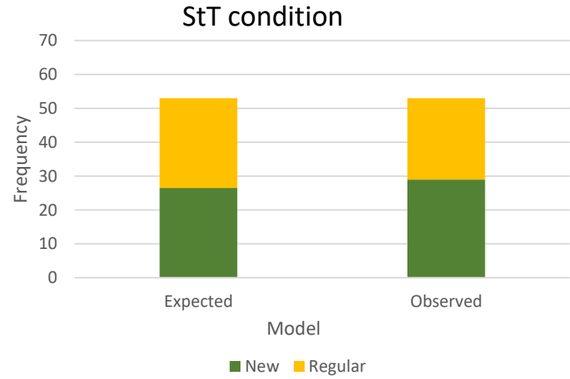

Figure 3. Frequencies of choosing the "new" or regular product for the expected model based on chance and the observed model in the StT condition.

Social desirability

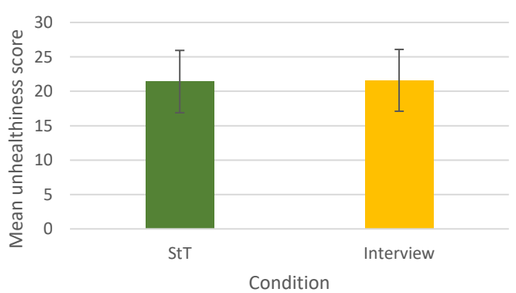

Figure 4. Mean unhealthiness scores for the StT and interview condition.

Conclusion

- **Acquiescence bias:** significantly higher satisfaction scores in the interview condition compared to the StT condition. Only seen for the "new" product.
- **Health claim:** significantly higher satisfaction scores for the "new" product compared to the regular product. Only seen in the interview condition.
- **Preference:** significantly higher preference of the "new" product in the interview condition.
- **Social desirability:** no significant differences between conditions, maybe due to carry-over effect.

The use of a human interviewer can create an interviewer bias. Using Speech-to-Text and Text-to-Speech techniques can limit this bias.