

COMPETITION IN THE ACOUSTIC ENCODING OF EMOTIONAL SPEECH

FRANK EISNER

*Adaptive Listening group, Max Planck Institute for Psycholinguistics, Wundtlaan 1,
6525XD Nijmegen, The Netherlands*

1. Introduction

Speech conveys not only linguistic meaning but also paralinguistic information, such as features of the speaker's social background, physiology, and emotional state. Linguistic and paralinguistic information is encoded in speech by using largely the same vocal apparatus and both are transmitted simultaneously in the acoustic signal, drawing on a limited set of acoustic cues. How this simultaneous encoding is achieved, how the different types of information are disentangled by the listener, and how much they interfere with one another is presently not well understood. Previous research has highlighted the importance of acoustic source and filter cues for emotion and linguistic encoding respectively, which may suggest that the two types of information are encoded independently of each other. However, those lines of investigation have been almost completely disconnected (Murray & Arnott, 1993).

2. Acoustic source vs filter cues

Source and filter components of the speech signal reflect distinct underlying articulatory mechanisms (Fant, 1960). The acoustic source originates at the level of the larynx through the quasi-periodic opening and closing of the vocal folds. Perceptual correlates of changing parameters in this vocal fold vibration such as jitter, shimmer, and especially its frequency (f_0) are important for emotional inflection in speech (Patel, Scherer, Björkner, & Sundberg, 2011). This source excitation is shaped by the resonances of the supralaryngeal vocal tract, acting as a dynamic filter by changing the constellation of the articulators. The identity of speech segments is largely determined by the filter (Nearey, 1989), but the importance of filter cues for emotion recognition has been considered to a much lesser extent in the literature.

3. Factorial approach

We investigated the relative contributions of these cues directly by using a vocoder technique to manipulate acoustic source and filter information independently (Eisner, Sauter, Hunt, Rosen, & Scott, n.d.). The results of a perception experiment showed that the two types of acoustic information have asymmetrical contributions to the recognition of emotional and linguistic content. An impairment of emotion recognition was found when either source or filter information was reduced. Source and filter channels were furthermore found to interact in emotion recognition such that the information carried by the filter was more beneficial in combination with original source information than without it. Comprehension of linguistic content, in contrast, was unaffected by altered source cues and relied only on filter cues.

4. Conclusions

Unlike nonverbal emotional vocalizations such as laughter or crying, paralinguistic expression of emotion competes with linguistic content in the acoustic filter channel, and this competition can be expected to further extend into the source channel in tonal languages. Independent manipulation of source and filter cues enables systematic investigations into how different types of information are encoded in speech and how they interact with each other.

Acknowledgements

This work was supported by Netherlands Organisation for Scientific Research and the Wellcome Trust.

References

- Eisner, F., Sauter, D. A., Hunt, H., Rosen, S., & Scott, S. K. (n.d.). Emotional expression in speech encoded by a combination of acoustic source and filter cues. Manuscript in preparation.
- Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.
- Murray, I., & Arnott, J. (1993). Toward the simulation of emotion in synthetic speech - a review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93(2), 1097–1108.
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, 85, 2088–2113.
- Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological psychology*, 87(1), 93–98.