Part IV

Free Talk Session 1

CONVERSATIONAL STYLES OF JAPANESE- AND ENGLISH-SPEAKING CHILDREN AND PARENTS

Yuko Yamashita
Department of English Communication, College of Engineering
Shibaura Institute of Technology, Tokyo, Japan

David Hirsh
School of Education and Social Work, University of Sydney, Sydney, Australia

The present study aimed at exploring the conversational speech style between parents and toddlers from English- and Japanese-speaking families. The study focused on (1) Children's responses toward parents and parents' responses toward children, (2) speech pauses, and (3) lexical use.

In study 1, the responses were classified into three categories: non-lexical backchannels, phrasal backchannels, and repetition. The results showed that the average ratio of overall backchannels and repetitions produced by parents was quite similar in both languages and much greater than that produced by children in the same languages. Among Japanese-speaking parents, non-lexical backchannels and repetitions were preferred to phrasal backchannels, while among English-speaking parents, non-lexical backchannels were most frequently used. With Japanese-speaking parents, almost half of the repetitions were exact in nature. They frequently repeated what a child had said and added the sentence-final particle "ne" or content words.

In study 2, all silent intervals (i.e., pauses) were extracted from each recording. We divided the pause into two types: intra-turn pauses, which occur within parents' or children's utterances, and inter-turn pauses, which occur between the turns of children and parents. Both intra-turn pauses and inter-turn pauses ranged from 0.1 up to 4 s. The results showed that all intra-turn pauses were less than 3 s for Japanese and English parents and 4 s for Japanese and English children. Similarly, all inter-turn pauses at turn transition from children to parents were less than 3 s, and 4 s for all pauses at turn transition from parents to children.

In study 3, all nouns and verbs produced by parents and children were coded and classified into five categories. This study followed the category of nouns and verbs proposed by Choi (1999). Nouns were categorized into object nouns, which referred to animate and inanimate entities, and non-object nouns, which included abstract nouns, locative nouns, and words that described activities or states (Bloom, Tinker & Margulis, 1993; Choi, 1999). Verbs were further categorized into action verbs, stative verbs, and mental verbs. The results showed that nouns and verbs were equally used by Japanese and English parents, while nouns were preferred by Japanese and English children. Additionally, object nouns were more frequently used than non-object nouns, and action verbs were more frequently used than stative and mental verbs in all groups.

These findings are expected to be useful in understanding conversational styles in spoken communication between parents and their children.

AUDITORY OBJECTS REPRESENTED AS TEMPORAL, SPECTRAL, AND SPATIAL MODULATION DENSITY PATTERNS: A THEORETICAL STUDY

Pierre L. Divenyi
Center for Computer Research for Music and Acoustics, Stanford University, USA
<pdivenyi@ccrma.stanford.edu>

Abstract

Experimental studies, e.g., those done at Shamma's and Theunissen's laboratories showing fields of cortical responses to spectro-temporal modulation patterns elicited by meaningful sounds in animals, suggest that similar patterns also underlie auditory object formation by human listeners. The present study expands this representation by combining spectro-temporal modulations with modulations in the spatial dimension, thereby making it possible to completely account for the perception of auditory objects, as defined in the three domains of "what", "when", and "where." Modulation patterns put psychoacoustics on a new footing: Considering objects to be specific modulation patterns in the three domains gives a formal definition to Wundt's concept of aperception; specifying the modulation density along any one dimension reinterprets Fechner's definition of psychophysical discrimination; assessing density changes along all three modulation dimensions combined offers a new way of seeing how and why simultaneous objects perceptually fuse or separate.

A perceptual object was defined by Wilhelm Wundt (Wundt, 1874) as a stimulus that possesses specific values along a set of physical dimensions but one which also represents a recognizable physical object that is meaningful to the person exposed to it. In other words, it should be not only perceived but also *aperceived* by the observer—the object should pass from the sensory-perceptual to the cognitive domain in which it would be learned, recalled, compared to traces of other similar objects, evaluated in terms of its value, meaning, and utility, etc. In particular, auditory objects are important to consider because the acoustic world with its myriads of omnipresent sounds cannot be shut out by humans and by most animals. Simple as they may appear in comparison to the complexity of optical images, sound waves just as these latter ones carry the same basic environmental information regarding the source that emits the stimulus: its identity, its temporal structure, and its location. While these three cardinal properties can be extracted from a waveform, and are decoded by peripheral stages of the binaural auditory system, further processing is necessary to analyze the output of these stages, in order to reconstruct the object emitted by the acoustic source. This more central second stage of analysis can be construed as a process determining the information density along the cardinal dimensions of spectral content, temporal organization, and spatial location. Spectral and temporal information density in the spectral and temporal domains is a result obtained by spectro-temporal modulation transfer functions of peripheral auditory processing in mammals (Kowalski et al., 1996), birds (Singh and Theunissen, 2003), and humans (Elhilali et al., 2003; Elliott and Theunissen, 2009). A recent attempt has been made for extending modulation transfer function analysis to the spatial domain (Divenyi, 2016). The present paper represents an attempt to combine the modulation analysis on all three domains in a theoretical study.

Method

The repertoire of acoustic signals included seven sounds, four of which were synthesized AM signals, one was speech-spectrum noise, and two were sentences taken from publicly available spoken corpora, one by a male talker from the SPIN test list and one by a female talker from the Harvard-IEEE corpus. All signals were 1.7 seconds in length. Each sound was processed to present it at one of 12 azimuthal positions separated by 30 ° over the entire horizontal circle, using the CIPIC head-related transfer function (HRTF) database's KEMAR head recordings with a 2-m radius. From these signals with different azimuthal positions 43 test sounds were generated: eight with single sounds at different azimuths, 12 with two sounds, four with three sounds, and two with four sounds combined. In addition, seven sounds were generated with one or two sounds at two different azimuths and two different AM rates. Peripheral processing of the sounds was emulated (1) by passing them through a bank of ERB-spaced second-order filters followed by an ear-like amplitude compressor, resulting in a cochleagram (Lyon, 1982), and (2) by generating a human-like azimuthal percept based on interaural time difference (ITD) weighted with interaural level difference (ILD), as proposed by Stern and colleagues (Stern et al., 1988), represented both with respect to temporal fluctuations and with the spectral profile of the sound. The modulation processing performed at more central (most likely cortical) sites was implemented by computing two-dimensional Fourier transforms of the three pairings of the three dimensions: time vs. frequency, azimuth vs. time, and azimuth vs. frequency.

Results

The questions that the modulation spectra of the pairs of dimensions allowed to ask were numerous. First, what is the information density along each of the dimensions? To assess the overall magnitude of the density is like evaluating the smoothness of a terrain, for which kurtosis is a reasonable measure. Second, high-intensity areas projected both on the abscissa and on the ordinate of the modulation spectrum reveal at what modulations is the information along the dimension corresponding to the axis is dense. These dense areas may possibly be typical representatives of the auditory object in question. Thirdly, jointly looking at the modulation density of the azimuth and either of the other two dimensions can reveal effects of reverberation (when the azimuthal modulation indicates multiple active angles and when the temporal modulation indicates the presence of a secondary source at a high modulation frequency (i.e., brief delay). Also, analysis of spectro-temporal modulation density can tell if there are two distinct objects or two sounds fusing into a single object—a test of fusion or segregation of two sources, an open question for quite some time (Bregman and Rudnicky, 1975).

Discussion

Clearly, the model as discussed represents the first phase of an in-depth study. Ideally, modulation spectra along the three dimensions would be computed and represented simultaneously in a 3-D drawing. However, one would have to take into account the fact that auditory space is already a three-dimensional concept: azimuth, elevation, and distance of any source viewed from an egocentric point, and such a three-dimensional dimension would make it hard to add it to the dimensions of time and spectrum. Nevertheless, the number of important psychophysical questions that could be treated with such a model may make it worth the effort. Time will tell.

Acknowledgments

The author feels greatly indebted to Drs. Frédéric Theunissen and Richard Stern for discussions about the ideas expressed in this paper.

References

Bregman, A. S., & Rudnicky, A. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance,* 1, 263–267.

Divenyi, P. L. (2016). Spatial modulation: Hearing the environment. Paper presented at the 22nd International Congress of Acoustics, Buenos Aires.

Elhilali, M., Chi, T., & Shamma, S. (2003). Intelligibility and the spectrotemporal representation of speech in the auditory cortex. *Speech Communication,* 41, 331–348.

Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.,* 5(3), e1000302. 10.1371/journal.pcbi.1000302.

Kowalski, N., Depireux, D. A., & Shamma, S. A. (1996). Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J. Neurophysiol.,* 76(5), 3503-3523.

Lyon, R. F. (1982). A computational model of filtering, detection and compression in the cochlea. Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Paris.

Singh, N. C., & Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *The Journal of the Acoustical Society of America,* 114(6), 3394–3411. doi:http://dx.doi.org/10.1121/1.1624067.

Stern, R. M., Zeiberg, A. S., & Trahiotis, C. (1988). Lateralization of complex binaural stimuli: A weighted-image model. *The Journal of the Acoustical Society of America,* 84(1), 156–165.

Wundt, W. M. (1874). *Grundzüge der physiologischen Psychologie* [Fundamentals of physiological psychology]. Leipzig: W. Engelmann.