Sensitive experimental techniques for the study of sound segregation

Alain de Cheveigné (CNRS/Université Paris 7),
Hideki Kawahara, **Minoru Tsuzaki** and **Kiyoaki Aikawa**
(ATR Human Information Processing Labs).

Introduction

Segregation phenomena are often studied using the "double-vowel" identification task: subjects are presented with stimuli containing two vowels, and requested to identify both of them [1]. The number of trials on which both vowels are correctly identified (*combination-correct* score) is used as a measure of segregation. This technique has several drawbacks: a) The task is often too easy and may lead to near-perfect identification in all conditions (ceiling effect); b) Since the subject must always answer two vowels, cues that signal the *multiplicity* of sources are ignored; c) The combination-correct score does not reveal eventual asymmetries between vowels within a pair.

This paper reports several modifications of that paradigm: an inter-vowel level mismatch was introduced to reduce ceiling effects, subjects were allowed to answer one or two vowels, and identification was scored separately for each vowel in a pair (*constituent-correct* score) [3,4].

Methods

Subjects were presented with stimuli that contained either one or two vowels, and were requested to give either one or two responses, rather than two as in the classic task. Vowels were synthetic Japanese vowels (/a/,/i/,/u/,/e/ and /o/), synthesized at F_0s of 125 and 132.5 Hz [2]. They were mixed to obtain double vowels with ΔF_0s of 0 and 6%. The relative level between vowels was -20 dB, -10 dB, 0 dB, 10 dB or 20 dB. Stimuli were 200 ms in duration, with 20 ms raised-cosine onset and offset ramps. All stimuli were set to the same RMS level before presentation over earphones, at a sound pressure level of 63-70 dBA. A complete stimulus set contained 840 stimuli (240 single and 600 double vowels) in random order. A session typically lasted 1-2 hours. Each subject performed 5 sessions, on separate days. There were six subjects, all native speakers of Japanese. For double vowel stimuli, each response was scored twice, once for each constituent vowel, to obtain constituent correct identification rates. The number of vowels (1 or 2) responded for each stimulus was also recorded.

Results

The number of vowels responded is plotted in Fig. 1 (top). At unison, subjects tended to hear two vowels when the constituents had the same level (0 dB), and one vowel when either vowel dominated. At ΔF_0=6% the pattern was similar, but the number of vowels responded was larger. Constituent-correct identification rates are plotted in Fig. 1 (bottom). Identification rates improved with target level, and were better when there was a ΔF_0 of 6% than at unison.

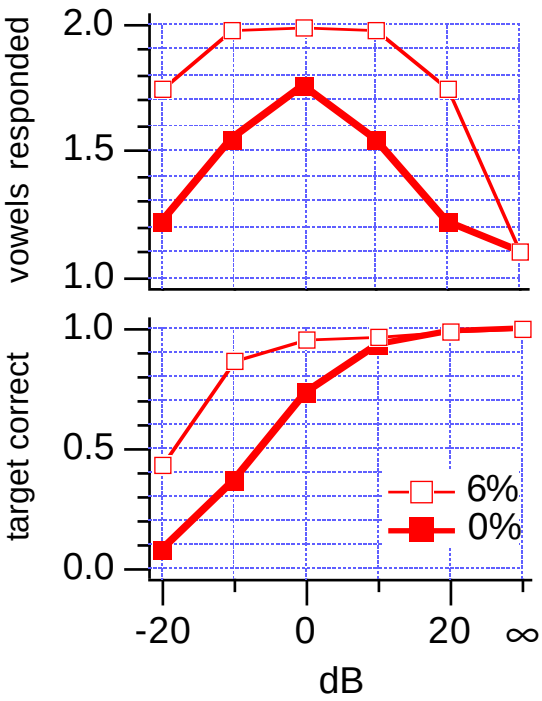

Fig. 1. Top: number of responses for each double-vowel stimulus. Bottom: target identification rate. Abscissa: level of target relative to ground. Rightmost point is for single vowels.

"Best" level mismatch

The ΔF_0 effect was largest at -10 dB, suggesting that this level mismatch effectively reduced ceiling effects. However, the increase in effect size would be useless if *variability* also increased in the same proportion. Fig. 2

shows that this was not the case: the ratio of the ΔF_0 effect (difference in rate between 0 and 6%) to its standard deviation calculated over subject, session and vowel pair, was also largest at -10 dB.

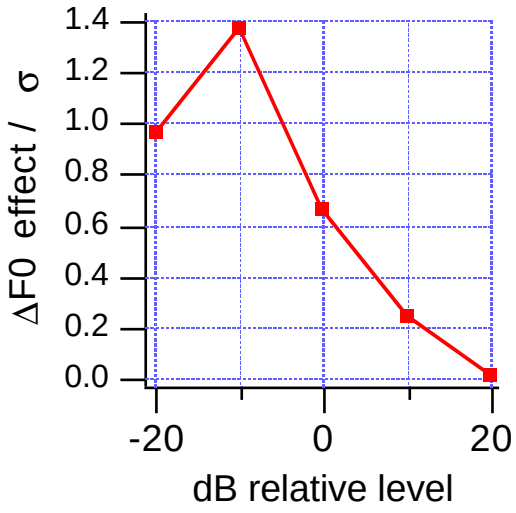

Fig. 2. Ratio between effect size and effect standard deviation.

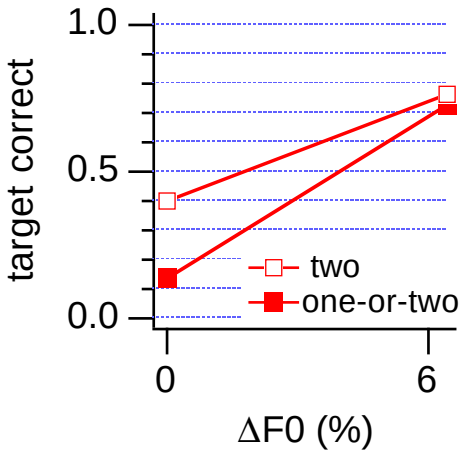

Fig. 3. Identification rate as a function of ΔF_0, for both tasks.

For experiments that involve a small perturbation from a baseline condition, it is important that baseline identification be neither too high (to avoid ceiling effects), nor too low (to avoid floor effects and subject frustration). For a baseline condition of ΔF_0=6%, a relative target level of about -15 dB should give an identification rate of about 70%. For a baseline condition of ΔF_0=0, a level of -5 dB would be more appropriate. The best level to avoid ceiling effects depends on the particular experiment planned.

One-or-two response task

At 0 dB relative level, the ΔF_0 effect we found was larger than effects previously reported: the combination-correct rate increased from 50 % (at unisson) to 90% (at ΔF_0=6%), whereas previous studies reported effect sizes of 15 to 30% [4]. Our larger effects were presumably due to the one-or-two response task we used.

In order to quantify the difference between tasks, we performed two new experiments. In the first, double vowel stimuli were constructed with a target level of -15 dB and ΔF_0s of 0 and 6.45%. The subjects were allowed to answer one or two vowels. In the second experiment, stimuli and subjects were the same, but subjects had to respond *two* vowels for every stimulus, whether or not they heard them. Results for both experiments are plotted in Fig. 3. Identification was better when subjects were forced to respond two vowels, but the increase was greatest at unisson. The consequence was a smaller effect size.

Conclusion

Reducing the level of one vowel relative to the other avoids ceiling effects. Allowing the subjects to answer *one or two* vowels makes the task easier, gives larger effects, and produces a measure of segregation: the number of vowels responded. Together these modifications make the double-vowel identification paradigm more sensitive.

Acknowlegments

This research was conducted under a collaboration agreement between ATR Human Information Processing Laboratories and the Centre National de la Recherche Scientifique (CNRS). Cécile Marin, Steve McAdams and Jean Laroche participated in the preparation. Hiroaki Kato and Ikuyo Masuda contributed useful ideas and advice. Rieko Kubo supervised the experiments. Thanks to John Culling, of the Nottingham Institute of Hearing Research for the synthesis software.

Bibliography

[1] Scheffers, M. T. M. (1983). "Sifting vowels" , unpublished doctoral thesis, Gröningen.

[2] Hirahara, T. and Kato, H. (1992). "The effect of F_0 on vowel identification", in Y. Tohkura, E. Vatikiotis-Bateson and Y. Sagisaka (Eds.), "Speech perception, production and linguistic structure", Tokyo: Ohmsha, 89-112.

[3] Lea, A. (1992). "Auditory models of vowel perception.", unpublished doctoral thesis, Nottingham University.

[4] de Cheveigné, A., McAdams, S., Laroche, J. and Rosenberg, M. (1995). "Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement." J. Acoust. Soc. Am. 97, 3736-3748.