# FORMANT BANDWIDTH AFFECTS THE IDENTIFICATION OF COMPETING VOWELS

Alain de Cheveigné
*CNRS - IRCAM, France, and ATR-HIP, Japan*

**ABSTRACT**

Formant bandwidth is known to have little effect on vowel quality. This paper shows that it has a strong effect on mutual masking between vowels. Subjects presented with stimuli consisting of pairs of synthetic vowels were requested to report one or two vowels for each stimulus. Identification rates were calculated independently for both vowels in the stimulus. Vowels had either the same or different fundamental frequencies. Their RMS amplitudes differed by 5, 15 or 25 dB. Formant bandwidth of each vowel was either twice or half its standard value. Identification of a target vowel was best when: (1) its RMS amplitude exceeded that of its competitor, (2) its formants were *narrow*, (3) formants of the competitor were *wide*, and (4) F0s were different. These effects were approximately orthogonal. A narrow-bandwidth voice is thus more resistant to masking, and a stronger masker, than a wide-formant vowel.

## 1. INTRODUCTION

Formant bandwidth is known to have little effect on the quality or intelligibility of isolated vowels [7,8]. However, if two vowels are in competition, as when two people speak at the same time, one can imagine that formant bandwidth might affect identification in several ways. For a given RMS amplitude, a formant attains locally a higher spectrum level if it is narrow than wide (Fig. 1), so a narrow-formant vowel might be more resistant to noise. On the other hand, interformant valleys are deeper when formants are narrow than wide, which might allow a competitor's formant peaks to emerge more easily. A narrow-formant vowel might thus be a less severe masker. There is therefore ample reason to suspect that formant bandwidth might affect identification of vowels *in competition*.

## 2. METHODS

The general methods are described in detail in [2,3]. In brief, stimuli were "double vowels" obtained by adding waveforms of two single vowels with amplitude ratios ranging from 5 to 25 dB in 10 dB steps. Single vowels were 5-formant synthetic Japanese vowels (/a/, /e/, /i/, /o/, /u/), with formant bandwidths one half or twice "normal", and fundamental frequencies ($F_0$) of either 124 or 132 Hz, allowing $F_0$ differences ($\Delta F_0$) of 0 and 6%. Formant frequencies were taken from [6], and "normal bandwidths" from [1]. Single vowels were synthesized with 270 ms durations, including 20 ms raised-cosine onsets and offsets. They had a "random" starting phase spectrum, the same for all vowels.

Single-vowels waveforms were scaled to a standard RMS amplitude after synthesis. To obtain a double vowel, two vowels were paired, one was scaled by a factor (5, 15 or 25 dB), both were added, and the sum was scaled to a standard RMS amplitude. The stimulus set included (20 pairs) x (3 amplitude ratios) x (2 $\Delta F_0$s) x (2 F0 orders) x (4 bandwidth combinations) = 960 stimuli. Stimuli were presented diotically via earphones. Sound pressure level varied between 63 and 70 dB(A) according to the stimulus

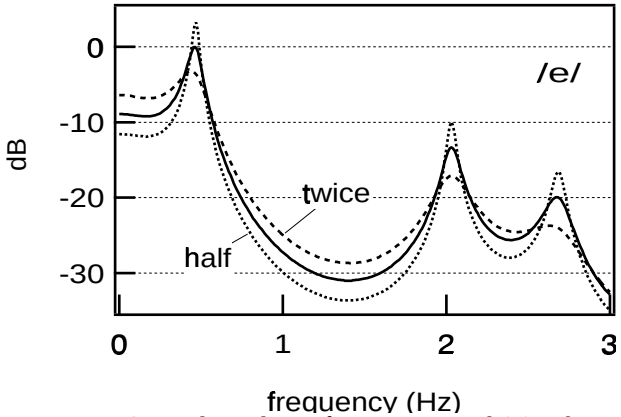


Figure 1. Spectral envelope of Japanese vowel /e/ with bandwidths that are normal (full line) or either half or twice normal (dotted lines).

Subjects (Japanese students, 8 male, 7 female, aged 18 to 22) were told that the stimulus set contained both double and single vowels (true for other experiments in the same series, but not this one) and requested to report *one or two* vowels for each stimulus, at will. Identification rates were measured independently for both vowels. A vowel was deemed identified if its name figured in the response for that stimulus. Rates were recorded as a function of *target bandwidth*, *competitor bandwidth*, their *amplitude ratio*, and the $\Delta F_0$. The number of vowels reported per stimulus was also recorded, but it is not described here.

| | /a/ | /e/ | /i/ | /o/ | /u/ | BW |
|---|---|---|---|---|---|---|
| F1 | 750 | 469 | 281 | 468 | 312 | 90 |
| F2 | 1187 | 2031 | 2281 | 781 | 1219 | 110 |
| F3 | 2595 | 2687 | 3187 | 2656 | 2469 | 170 |
| F4 | 3781 | 3375 | 3781 | 3281 | 3406 | 250 |
| F5 | 4200 | 4200 | 4200 | 4200 | 4200 | 300 |

Table 1. Formant frequencies [6] and bandwidths [1] of vowels used in the experiment.

## 3. RESULTS

A statistical analysis was performed independently at each amplitude ratio by means of repeated-measures ANOVAs with

factors ($\Delta F_0$ = 0, 6%) x (target bandwidth = half, twice) x (competitor bandwidth = half, twice). Only effects significant at p=0.05 are discussed here.

### 3.1 Formant bandwidth

Figure 2 shows the identification rate as a function of target and competitor bandwidth at each amplitude ratio. Dotted lines connect conditions that differ by the target's bandwidth. They all have a negative slope: all else being equal, *identification was better for narrow- than for wide-formant targets*. Full lines connect conditions that differ by the competitor's bandwidth. They also all have a negative slope: *identification was better for vowels in competition with a wide- than a narrow-formant vowel*. At -5 dB the lines form a parallelogram, indicating that the two effects are *independent*. At other ratios the shape is less regular, but this can be interpreted as the result of a sigmoid distortion reflecting ceiling and floor effects. Target and competitor bandwidth effects have similar sizes, with the result that identification is the same in the n/n and w/w conditions (except at -25 dB). A similar pattern prevailed at $\Delta F_0$=6% (not shown), with overall higher rates.

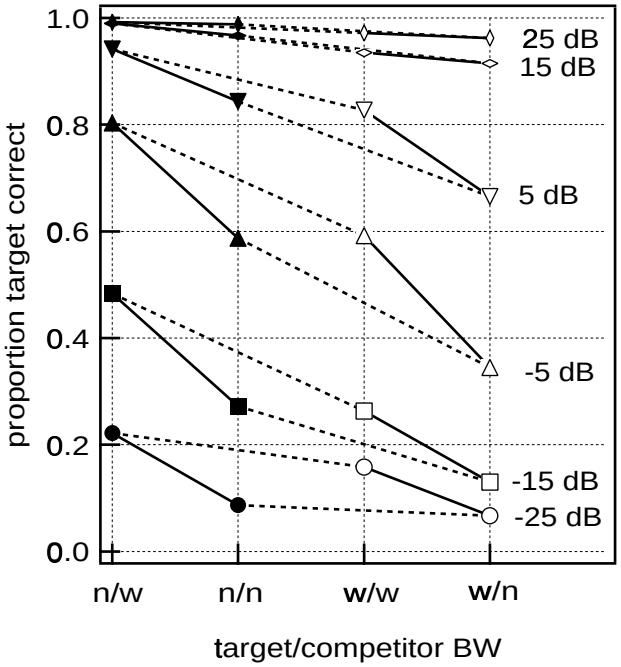


Figure 2. Identification rate as a function of formant bandwidth (target/competitor), at each amplitude ratio, at $\Delta F_0$=0.

Comparing effects of bandwidth at constant ratio with those of ratio at constant bandwidth (Fig. 2), it appears that a 4-fold change in bandwidth has an effect similar (in general slightly smaller) to that of a 10 dB change in ratio. Referring back to Fig. 1, a 4-fold reduction of bandwidth increases the formant peak amplitude by about 6 dB. One could thus explain target bandwidth effects by assuming (a) that narrowing a formant at constant RMS boosts its peak amplitude, and (b) that perceptual salience depends on the amplitude localized at *formant peaks*, rather than averaged over wider ranges, or the whole spectrum. The similar (and orthogonal) effects of competitor bandwidth suggest an analogous explanation: a vowel's masking power depends on the amplitude of its formants at their peak, rather than its RMS amplitude, or the spectrum level in the vicinity of the target's formants (as was hypothesized in the Introduction).

This explanation is problematic in at least two ways. First, it assumes precise sampling of the envelope amplitude at a formant peak, which is hard to reconcile with the smoothing steps that are often included in models of vowel perception. It also ignores the difficulty of estimating the amplitude of a narrow peak (45-55 Hz in the narrow condition) sparsely sampled by harmonics spaced at 124 or 132 Hz intervals [4]. Second, in the case of competitor bandwidth effects, the explanation supposes direct competition between vowels: the target is masked in proportion to the salience of the formant cues belonging to the competitor. This is in contradiction with conclusions of a previous study [5] that found evidence that cues to both vowels could coexist, as long as their own salience was not affected. This contradiction reveals the limits of qualitative interpretations in terms of "feature salience", as performed in that study and here. Its resolution probably requires simulation with computational models of concurrent vowel perception.

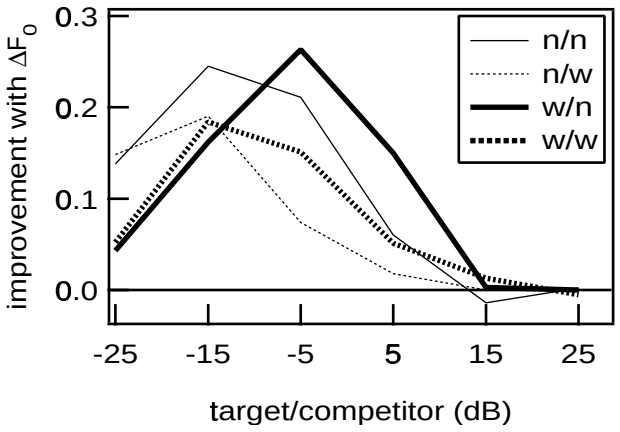


Figure 3. Improvement in target identification provided by a 6% $\Delta F_0$, as a function of target/competitor ratio, for each bandwidth condition.

### 3.2 $\Delta F_0$ effects

Identification was better at $\Delta F_0$=6%, as observed in many previous studies (eg. [1]). The improvement [i($\Delta F_0$=6%) - i($\Delta F_0$=0)] is plotted in Fig. 3 as a function of amplitude ratio, for each bandwidth condition. The four lines have similar shapes. Their downward slope at large ratios reflects a ceiling effect: $\Delta F_0$ is of no help if identification is already perfect. Consistent with this idea, the right-hand edges of the four lines are staggered in inverse order relative to the rates plotted in Fig. 2 (In Fig. 3 $\Delta F_0$ effects extend to highest ratios for w/n, for which identification rates were lowest at

$\Delta F_0$=0 in Fig. 2). The drop-off on the left-hand side reflects the breakdown of segregation mechanisms at low target amplitudes.

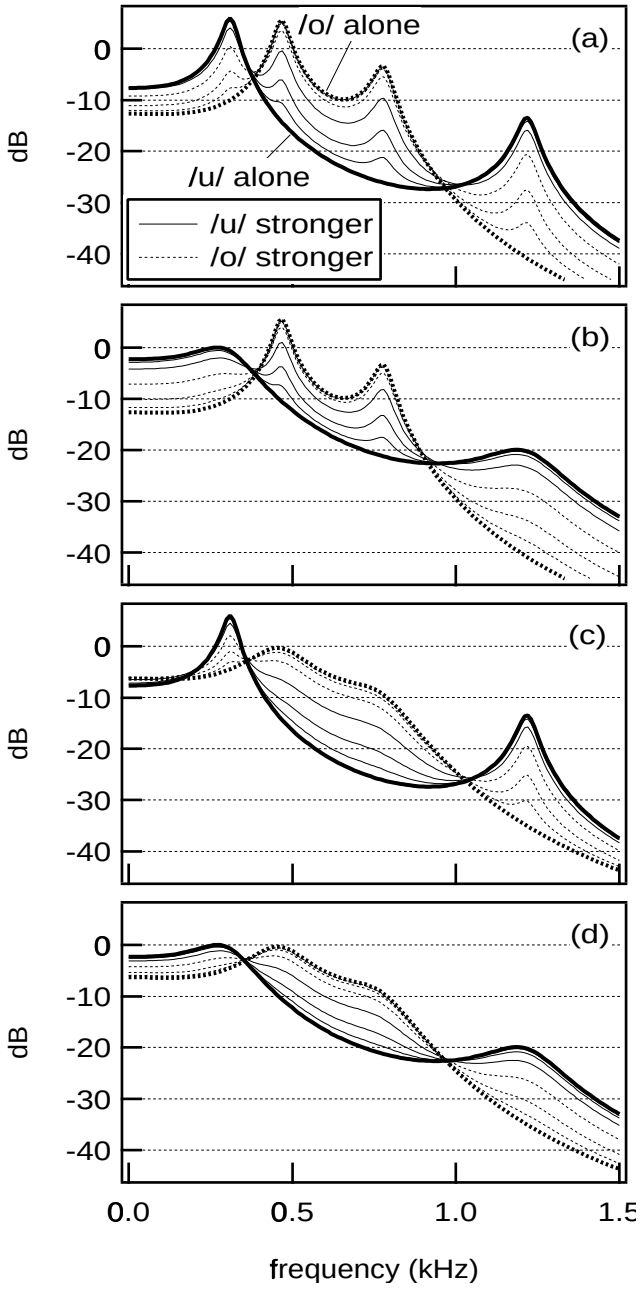


Figure 4. Spectral envelopes of stimuli for vowel pair /o+u/. Each panel is for a different bandwidth condition (twice or one half normal bandwidth). Within a panel, thick lines represent single vowels (not used as stimuli). Thin lines represent double vowels dominated by /u/ (full lines) or by /o/ (dotted lines).

It is interesting to note in this respect that $\Delta F_0$ effects at -25 dB were larger for targets with formants that were narrow rather than wide. This might be an effect of the greater amplitude at the peak of their formants. Conversely, overall across ratios it appears that the magnitude of the $\Delta F_0$ effect is larger for narrow- than for wide-formant competitors. It would seem that the masking power of a narrow-formant vowel surrenders more easily to a $\Delta F_0$ difference.

### 3.3 Pairwise effects

Subjects' responses may also be analyzed separately for each vowel pair. The appeal of such an analysis is that responses for each condition may then be compared with the spectrum of the stimulus for the same condition. The diversity of conditions (6 amplitude ratios) x (4 bandwidth combinations) for each pair allows a fine-grained analysis. The difficulty with this proposition is the large volume of data for the 20 target/competitor pairs, and the relatively small number of trials for each data point (30), that limits the reliability of effects observed [5]. Data for the pair /o+u/ will be presented in detail, to illustrate the potential of such an analysis, and also its limitations.

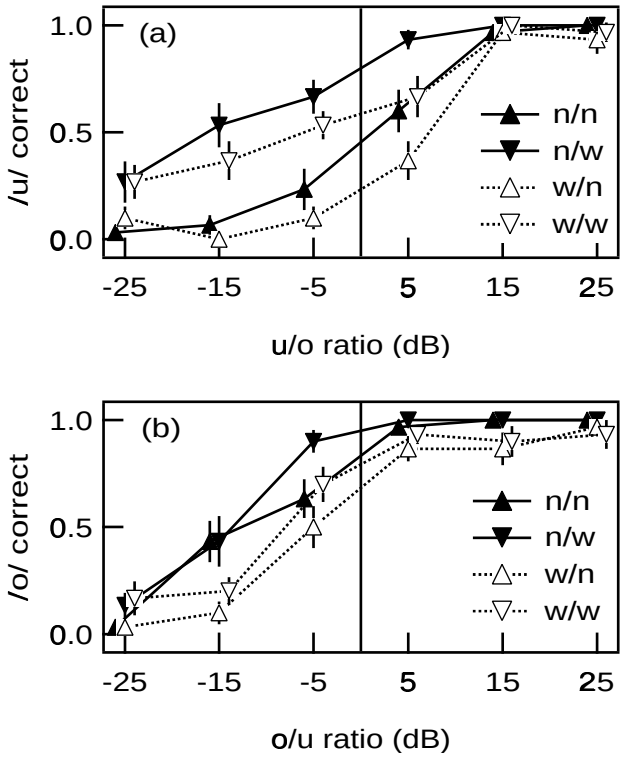


Figure 5. Target identification rate as a function of the target/competitor amplitude ratio, for each bandwidth condition, for vowel pair /o+u/. Top: u/o. Bottom: o/u. Vertical bars represent one standard error of the mean.

All single vowels had the same starting phase spectrum (Methods), and partials of same rank therefore summed in phase. There is thus no need to consider phase-dependent vector summation: the spectral envelope of a double vowel is simply the sum of the envelopes of its constituents, each

weighted with the appropriate scaling factor. As the amplitude ratio varies between -25 dB and 25 dB, the envelope of the stimulus transmutes between a shape similar to the first vowel and a shape similar to the second. This is illustrated in Figure 3 (the graph is restricted to formants F1 and F2 which are most important for identification).

A vowel's formants are *visually* more prominent when their bandwidths are narrow rather than wide. Compare for example formants of /o/ between Fig. 3 (a) and (c), or between (b) and (d). This is coherent with the target bandwidth effects observed.

The effect of the competitor's formant bandwidths on the prominence of the target's formants is harder to judge. On one hand, inter-formant valleys are deeper for a narrow-formant competitor (as suggested in the Introduction). On the other hand, a wide-formant competitor offers a flatter and lower "spectral context" for the target's formants. As argued in Sect. 3.1, the experimental data suggest that the latter factor is determinant.

Examination of the pair-specific data plotted in Fig. 5 reveals some important differences with the data averaged over pairs (Fig. 2). To start with, effects of target and competitor bandwidth no longer have the same size: for the vowel pair u/o, the n/n and w/w conditions are equivalent at 5 dB and above, but they differ below. For o/u they differ at -15 dB. Identification of /u/ is strongly affected by the formant bandwidths of its competitor, and less by its own. The opposite is true of /o/, at least at o/u ratio = -15 dB. The simple relation between magnitudes of bandwidth effects and amplitude ratio effects also no longer holds. This in turn casts doubt on the simple account of bandwidth effects in terms of amplitude-at-formant peak, outlined in Sect. 3.1.

Discussion of such pair-specific effects is limited by the complexity of the data, pointed out earlier, and also by the uncertain relevance of the visual prominence of spectral envelope cues to the quality of information available to the auditory system. The next step in exploring such data should involve simulations with computational models of auditory processing (including masking) and concurrent vowel identification, for which such a data set would constitute an excellent test bed.

## 4. DISCUSSION

An unexpected outcome of the experiment was that sharpening a competitor's formants increased, rather than decreased, its masking power, despite the fact that interformant valleys are deeper for narrow formant vowels. Target identification seems to depend on prominence of its formants relative to those of the competitor, rather than relative to the local spectrum level. Such a direct competition is inconsistent with data gathered in a previous study [5]. More research is required to resolve this question.

Formant bandwidth reflects losses within the vocal tract, for example damping during the open glottis phase. Such losses may vary with changes in phonation style, in particular stress. One could speculate that an effect of vocal stress might be to reduce acoustic losses (by shortening the glottal closed phase, and possibly stiffening tissues bounding the vocal tract), and thus give the speaker's voice a *competitive* edge, enhancing both its masking power and its resistance to masking. If so, it might give the speaker an advantage in competitive social situations.

## 5. CONCLUSION

Formant bandwidth affects the identification of vowels in competition with other vowels. At constant RMS amplitude, identification of a vowel is enhanced by sharpening its formants, or widening those of its competitor. Effects of target and competitor bandwidth are approximately independent, and independent with those of amplitude ratio and $\Delta F_0$. The effect of a 4-fold change in target or competitor bandwidth is roughly of the same order of that of a 6% $\Delta F_0$, or a 10 dB change in target/competitor amplitude ratio.


### ACKNOWLEDGEMENTS

This work was performed within an agreement between ATR Human Information Processing Research Labs, CNRS, and University Paris 7. Rieko Kubo, of ATR ran the experiments, and Hideki Kawahara contributed valuable advice. The stimuli were synthesized with John Culling's |WAVE software.


### REFERENCES


[1] Assmann, P. F., and Summerfield, Q. (1990). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. 88, 680-697.

[2] de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (1997). "Concurrent vowel identification I: Effects of relative level and F0 difference," J. Acoust. Soc. Am. 101, 2839-2847.

[3] de Cheveigné, A. (1997), "Ten experiments in concurrent vowel segregation," ATR Human Information Processing Research Labs technical report, TR-H-217.

[4] de Cheveigné, A., and Kawahara, H. (1999). "Missing data model of vowel perception," J. Acoust. Soc. Am. (accepted for publication)

[5] de Cheveigné, A. (1999). "Vowel-specific effects in concurrent vowel identification," J. Acoust. Soc. Am. (accepted for publication)

[6] Hirahara, T., and Kato, H. (1992). "The effect of F0 on vowel identification," in "Speech perception, production and linguistic structure," Edited by Y. Tohkura, E. Vatikiotis-Bateson and Y. Sagisaka, Tokyo, Ohmsha, 89-112.

[7] Klatt, D. H. (1982). "Prediction of perceived phonetic distance from critical-band spectra: a first step," Proc. IEEE ICASSP, 1278-1281.

[8] Rosner, B. S., and Pickering, J. B. (1994). "Vowel perception and production," Oxford, Oxford University Press.