A Perceptual Study of the Influence of Pitch on the Intelligibility of Sung Vowels

Nicole Scotto Di Carlo, Aline Germain

Institut de Phonétique d'Aix-en-Provence, U·A·261, CNRS, France

Abstract. A perceptual study of the influence of pitch on the intelligibility of vowels was carried out using a corpus containing all 15 French vowels sung by a professional soprano across her entire voice range. Four untrained subjects underwent identification tests whose results show statistically that vowel intelligibility is inversely proportional to pitch. A perceptual analysis based on confusion matrices revealed that intelligibility drops rapidly starting at the middle register. Classification of the confusions showed that incorrectly identified vowels tend to be confused with [a], most certainly because the shape of the vocal tract when in the upper register corresponds to that of the vowel [a].

Introduction

All opera lovers know how difficult it is to understand words that are sung. For the past 20 years or so, this phenomenon has aroused researchers' interest and has been the basis of various articles attempting to define the roles played by pitch, intensity, rate of production, and vibrato in the comprehension of vowels and consonants [Husson, 1957, 1958; Cornut and Lafon, 1960; Howie and Delattre, 1962; Scotto Di Carlo, 1972, 1978, 1981; Sundberg, 1975; Smith and Scott, 1980; Johansson et al., 1982; Germain and Séassau, 1982). We shall limit our study here to the influence of pitch on the intelligibility of French vowels in singing.

Experimental Procedure

Choosing the Subjects and Preparing the Corpus

Studies carried out by Strange et al. [1976], Janson [1980], Smith and Scott [1980], and Gottfried and Chew [1986] have shown that in speaking as well as in singing, isolated vowels are not perceived as well as vowels in a CVC context. In order to bring together the conditions under which intelligibility is most highly perturbed, and to determine how well the various vocalic cues resist pitch distortion, we chose a corpus of isolated vowels sung by a professional singer with a coloratura-soprano voice, the highest-pitched female voice. (A similar analysis is now being made on all other types of voices.)

Recording the Corpus

The recording was made in an anechoic chamber on a Nagra I single-track tape recorder equipped with a Philips LBB 9060 microphone. A professional singer was requested to sing the 15 French vowels, of approximately equal duration, across her entire voice range. Since the subject could not achieve the same pitch range for each vowel, the smallest range attained was used so that the corpus would be uniform and would contain 17 different pitch levels for each of the 15 vowels, that is, 255 items.

Setting up the Tests

The limits of each vowel on the tape were determined auditorily, the tape was cut at these points, and the pieces were assembled in three different orders, pre-established by a random number generation program, to form three test tapes. Stimuli were separated from each other by a 2.5-second pause, and a longer pause was inserted every 50 stimuli so that the subject could rest.

Perception Tests

Taking the Tests

Before each test, written instructions were given to the subject explaining that he was to write down on the chart provided the phonetic symbol for each of the vowels he heard. The phonetic symbols were reviewed before the test in order to avoid transcription errors.

The 4 subjects chosen were phoneticians but not musicians, and were not trained for this test. The three test tapes were presented in a different order to each subject so as to avoid any possible biasing of the final results due to sequence effects. It was suggested to us that the subjects listen to a short extract of an opera sung by the same person that had recorded the corpus in order to provide the subject with a reference framework (we are grateful to D. J. Hirst for his suggestion). Subjects could read the words to the opera while listening, which allowed them to become accustomed to the singer's vocalic system and to begin each test under identical conditions. Indeed, according to Janson [1980], the results of tests on phonemic identification are highly influenced by the linguistic environment to which the subject is referring during the tests. This explains why the same test, taken by the same subject, under the same conditions may give extremely different results each time taken. Imposing a linguistic environment at the beginning of each test provides short-term memory conditioning, and leads to more coherent results when tests are spread across time.

Statistical Processing of the Results

An analysis of variance (repeated measures design) was carried out based on a sampling of frequencies representing each of the five production modes [Hoc, 1983]. We make the distinction between 'register', which designates a part of the vocal scale (i.e. lower, lower middle, middle, upper middle, upper), and 'production mode', which designates the manner in which the sound is produced (i.e. chest voice, chest mid voice, mid voice, head mid voice, and head voice or falsetto). Indeed, the word 'register' is usually used to refer to both the parts of the musical scale and the way the sound is produced because it is generally accepted that there is exactly one register for each production mode (for instance, lower register/chest voice). The confusion of the two is an erroneous simplification of the actual situation, since it is possible to produce the notes of a given register in different production modes (for instance, a note in the middle register may be produced with any of the following three voice mechanisms: chest, middle, or head voice).

The experimental design was as follows:
R3*S4*V15*P5
where R = repetitions, S = subjects, V = vowels, P = production mode, R being the random factor. In a first analysis, R was the random factor. Values for F were as follows:

Percentage of identification: 65% Matrix No. 5
Number of correct responses: 118 Pitch: 330 Hz (E_4)

Vowels to be identified

Vowels perceived	i	e	ɛ	a	ɑ	y	ø	œ	u	o	ɔ	ɛ̃	œ̃	ɔ̃	ɑ̃
i	12														
e		12													
ɛ			1												
a				8									1		
ɑ				3	1		1				1		1	2	
y						12									
ø			4				8	3							
œ			2				2	2					1		
u									11						
o									1	12	2				
ɔ					1						9			1	
ɛ̃			3					4				12	2		
œ̃			2	1			1	3					7		
ɔ̃					1									0	1
ɑ̃					9									9	11

Fig. 1. Example of a confusion matrix [based on Miller and Nicely, 1955]. Number of occurrences of the stimulus/response pair is recorded in the corresponding box of the matrix. Example: The vowel [ɔ̃] was mistaken for [ɑ] twice at 330 Hz.

V: F (14–28) = 31.25, p <0.00001
P: F (4–8) = 102.43, p <0.00001
VP: F (56–112) = 6.28, p <0.00001
S: F (3–6) = 1.76, p <0.2549

In a second analysis, S was the random factor. Values for F were as follows:

V: F (14–42) = 10.85, p <0.00001
P: F (4–12) = 171.51, p <0.00001
VP: F (56–168) = 5.14, p <0.00001
R: F (2–6) = 0.028, p <0.97281

According to these analyses, the factors S and R are not significant, which shows that the across- and within-subject responses are homogeneous. However, the factors V, P, and the V–P interaction are highly significant. Therefore, subjects' responses are highly influenced both by the nature of the vowel and by the frequency at which it is sung, as initially hypothesized.

Perceptual Analysis

Analysis Using a Vowel Confusion Matrix

Method

The method used for analyzing the results is based on that of Miller and Nicely [1955]. The primary interest of this method is that it is not limited to determining the percentage of vowels identified correctly, which provides no information on the nature of the perceptual confusions, but allows for recognizing and explaining the types of errors that occur. This is done using confusion matrices made up of a finite number of features that lead to the establishment of the perceptual hierarchy of those features. The subjects' responses are totalled in the confusion matrices (one matrix per pitch level). The vowels to be identified are listed horizontally and the vowels that were perceived are indicated verti-

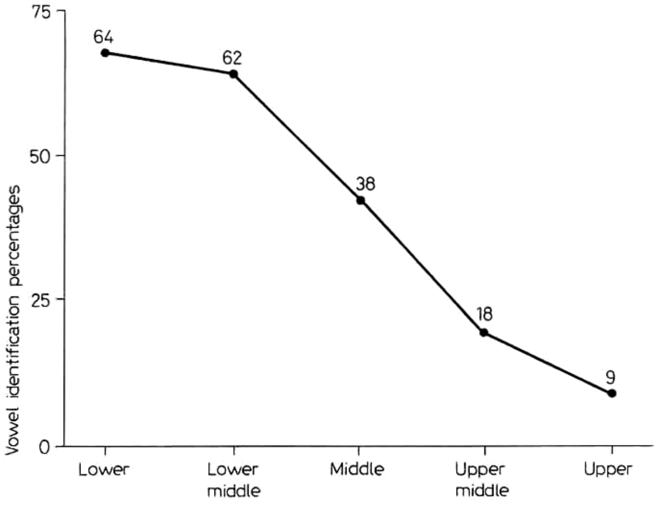

Fig. 2. Vowel identification. Vowel identification percentages are given for each register.

cally. Thus, each position within the matrix represents the number of occurrences of the stimulus/response pair. The number of correct responses is obtained by adding up the numbers on the main diagonal (fig. 1).

Results

The correct identification percentage for each matrix was found to be inversely proportional to pitch. This tendency is even more marked when the pitch levels are grouped into registers [the register limits were determined according to the method of Aristopoulos, 1983]. For all vowels, there is an abrupt decrease in intelligibility beginning with the middle register where the percentage of identification (when all vowels are taken together) falls to 38% and then to 9% in the upper register, whereas it is 64 and 62% in the lower and lower-middle registers (fig. 2). It should be noted, however, that when the singer is shifting modes, there is a significant decrease immediately followed by a sharp increase in the percentage of identification (table I).

Interpretation

When the singer reaches the upper limit of a register, she shifts modes by tuning the resonance cavities to the laryngeal sound [Tarneaud, 1961; Coffin, 1976; Dutoit-Marco, 1985]. The resulting adjustment of the bucco-pharyngeal cavity somehow neutralizes the modifications of the buccal cavity that are required for producing vowels. This explains why their intelligibility is sacrificed. As soon as the change in production mode is completed, however, the singer reestablishes the optimal phonatory position

Table I. Recognition percentages

Pitch hz	Note	Register	Production mode	Percentage of recognition
220	A_3	lower	chest voice	70
247	B_3			66
262	C_4			58*
294	D_4	lower middle	chest mid voice	65*
330	E_4			65
349	F_4			57*
392	G_4	middle	mid voice	65*
440	A_4			47
494	B_4			35
523	C_5			35
588	D_5			26
659	E_5			20*
698	F_5	upper middle	head mid voice	28*
784	G_5			15
880	A_5			13
988	B_5	upper	head voice	10
1,046	C_6			8

Percentages of recognition are given by register and production mode for each pitch level. Stars indicate changes in production mode that lead to significant modifications in recognition rate.

for the distinct vocalic timbres. In order to study the vocalic confusions as a function of the phonetic make-up of the vowels, we complemented the above analysis, based on the percentages of identification, with an analysis by feature.

Analysis by Feature

The vowels the singer was requested to sing were first categorized according to their distinctive physiological features, as follows: labiality (lip rounding), jaw opening, nasality, place of articulation, aperture (maximum degree of constriction of the vocal tract), and height of the tongue.

This classification, when applied to the 17 matrices, provided the percentage of correct identifications for each feature. The results for the first three features, grouped by register, are shown in figure 3.

Labiality

Results. Labiality is the feature that resists the least in the upper register. The percentage of labial vowels identified goes from 63 to 60% for the lower and lower-

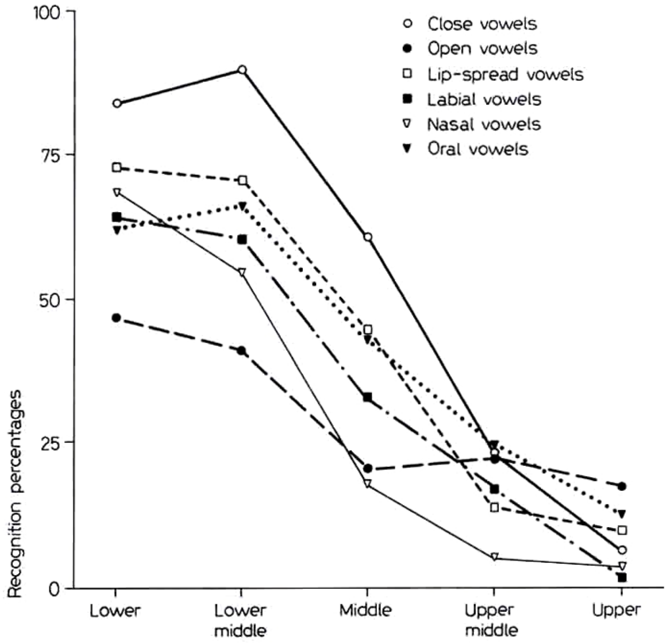

Fig. 3. Recognition percentages. Percentages of recognition are given for the main vowel features as a function of register.

middle registers, and to 33, 17, and finally 1.5% for the middle, upper-middle, and upper registers, respectively.

Interpretation. This is due to the fact that the production of notes located in the lower register is generally accompanied by a more or less prominent lip rounding designed to increase the volume of the buccal cavity and to amplify the low-frequency harmonics. On the other hand, the vocal technique used by our subject for the production of notes situated in the upper register is characterized by a lateral spreading of the lips aimed at compensating for the lowering of the jaw by decreasing the volume of the buccal cavity so as to amplify the high-frequency harmonics. This explains why lip-spread vowels such as [i] and [e] are relatively well perceived at higher pitch levels.

Jaw Opening

Results. The correct perception of close vowels only occurs when they are produced in the lower and lower-middle registers, and starts deteriorating in the middle register. Indeed, the percentage of identification of these vowels, which is 78% in the lower register and 88% in the lower-middle register, falls to 59% in the middle register and to 4% in the upper register. For open vowels, on the other hand, which are correctly perceived nearly half of the time in the lo-

Table II. Vowel confusion

Percentage out of total number of mistaken vowels	Mistaken for
34	[a]
8	[ɑ̃]
7	[œ]
7	[ɑ]
7	[o]
7	[ɔ]
6	[ø]
6	[ɛ̃]
6	[œ̃]
4	[e]
3	[ɛ]
2	[y]
2	[ɔ̃]
1	[i]
1	[u]

Line 1, for example, is read '34% of the vowels that were incorrectly identified were perceived as [a]'. For correct identification percentages, see table III.

wer and lower-middle registers (46 and 40%, respectively), the percentage of correct identification decreases slightly starting at the lower-middle register, and then remains stable all the way up to the upper register where it is 19%.

Interpretation. As numerous authors have shown, jaw opening is directly proportional to the fundamental frequency of the sound produced. That is, the higher the fundamental frequency, the greater the interincisor distance. It is thus easy to understand why it is difficult to identify close vowels in the upper-middle and upper registers, where the buccal opening is great.

Nasality

Results. The mean value for the percentage of nasal vowels properly identified when all frequencies are evaluated together is relatively low (29%). Within the 79% of incorrectly identified vowels, 53% were heard as oral vowels and 18% as nasal vowels of the wrong timbre.

Interpretation. Such a low recognition level for nasal vowels may be explained by the fact that singers avoid nasalizing for aesthetic reasons. It is indeed difficult to sustain a vowel without the occurrence of a dorso-uvular occlusion because the inertia of the soft palate rapidly leads to a purely nasal sound rather than to the oro-nasal sound characteristic of the nasal vowels. In order to delay the dorso-uvular occlusion process when producing nasal vowels, singers attack the corresponding oral vowel first, and nasalize afterwards.

The inertia of the velum during the emission of French sung nasals is currently being studied in a research project sponsored by the French Ministry of Culture. The statement made in our article has been shown to be true, but results have not yet been published.

Secondary Features

Results. The features representing place of articulation, aperture, and height of the tongue are not significant.

Interpretation. Indeed, singing requires both a vocal tract that is completely free of constriction and a high degree of flexibility in the lingual and bucco-facial muscles. The French vocalic system, however, is characterized both by the constriction of the vocal tract and by a great deal of muscular tension due to the high proportion of front vowels [Delattre, 1953]. A possible interpretation would be that the singer attempts to reduce this constriction and muscular ten-

Table III. Correct vowel identification

	i	e	ɛ	a	ɑ	y	ø	œ	u	o	ɔ	ɛ̃	œ̃	ɔ̃	ɑ̃
Lower															
A_3	1	0.91	0.50	0.50	0.41	1	0.66	0.50	0.91	1	0.41	0.83	0.75	0.25	0.91
B_3	1	1	0.25	0.58	0.25	0.91	0.83	0.41	0.16	1	0.75	0.41	0.58	0.83	0.91
C_4	0.58	1	0.33	0.41	0.33	0.33	0.58	0.58	0.25	1	0.66	0.91	0.50	0.33	0.91
Lower middle															
D_4	0.91	1	0.16	0.75	0.25	0.91	0.58	0.66	0.66	0.83	0.50	0.83	0.50	0.33	0.91
E_4	1	1	0.08	0.66	0.08	1	0.66	0.16	0.91	1	0.75	1	0.58	0	0.91
F_4	1	0.91	0	0.83	0.25	1	0.58	0.25	0.91	1	0.41	0.33	0.25	0.08	0.83
Middle															
G_4	1	0.83	0	0.75	0.16	1	0.50	0.08	0.58	0.83	0.08	0.41	0.58	0.83	0.41
A_4	1	1	0.16	0.83	0.08	1	0.41	0.08	0.66	0.50	0.08	0.16	0.16	0.41	0.50
B_4	0.58	0.91	0	0.66	0.08	1	0.33	0	0.91	0.33	0.08	0.08	0.08	0	0.25
C_5	0.83	0.75	0.33	0.50	0.25	0.91	0.25	0.16	0.66	0.25	0.16	0	0	0.08	0.16
D_5	0.50	0.41	0.08	0.41	0	0.91	0.08	0	0.66	0.33	0.25	0	0.08	0	0.16
E_5	0.58	0.58	0	0.66	0.25	0.25	0.08	0.08	0	0.16	0.16	0	0	0.08	0.16
Upper middle															
F_5	0.50	0.50	0	0.91	0.08	0.58	0.25	0.33	0.33	0.33	0.25	0	0	0	0.16
G_5	0.25	0.08	0	0.58	0.08	0.08	0.16	0.08	0.33	0.25	0.33	0	0.08	0	0
A_5	0.16	0.16	0.08	0.58	0	0	0.16	0.08	0.16	0.16	0.25	0	0.08	0.08	0.08
Upper															
D_5	0.25	0.16	0	0.75	0.16	0	0	0	0	0	0.08	0	0	0	0.08
C_6	0.16	0	0	0.91	0	0	0	0.08	0	0	0	0	0.08	0	0

Correct identification of each vowel over the 17 notes (and 5 registers). Recognition rate = total number of correct answers / n, where n = 12 (1 vowel × 4 subjects × 3 repetitions).

sion, both of which are incompatible with the requirements of singing, by centralizing the articulation of the vowels, that is, by underarticulating them.

Hierarchy of the Confusions

Results. Using the 17 matrices, the confusions were classified hierarchically. In the cases where the vowels were not properly identified, the subjects tended to confuse them with the open and central vowels, and in particular with [a] (tables II, III).

Interpretation

Confusion with [a] most likely occurs because the shape of the vocal tract when singing in the upper register (buccal cavity volume > pharyngeal cavity volume, tip of the tongue on the floor of the mouth) corresponds to that of the vowel [a]. From an acoustical point of view, the predominance of [a] may be explained, according to Howie and Delattre [1962], by the fact that whenever the fundamental frequency is greater than 750 Hz (the value of the first formant of [ɑ] according to Delattre), the ear perceives an intermediate formant located between the theoretical F_1 and F_2 of [ɑ].

Conclusion

As we have seen, the physiological movements aimed at reducing the constriction of the vocal tract and at lessening the degree of muscular tension caused by the vocalic anteriorization of the French language result in under-articulation [Scotto di Carlo, 1978]. This under-articulation is detrimental to the precision of the vocalic timbres which is an important factor of their intelligibility. In addition, the bucco-pharyngeal adjustments necessary to vocal production are not always compatible with the modifications of the buccal cavity that are required for phonemic articulation (in particular during mode shifts). When the former happen to coincide with the latter, the intelligibility of the vowel is preserved. This is the case, for example, for the labial vowels and the close vowels in the lower register, and for the open vowels and the lip-spreading vowels in the upper register. In the other cases, the degree to which intelligibility is sacrificed depends upon the extent to which the two types of adjustments are antagonistic.

Acknowledgements

We would like to thank Mme. Mady Mesplé, who was kind enough to allow us to record her voice, Christian Cavé, Daniel Hirst, and Yukihiro Nishinuma, CNRS researchers, for their advice concerning the perceptual part of our study, Henriette Séassau for her aid in setting up the tests, and Vivian Lamongie for the English version of this article. This study was sponsored by a grant from the French Ministry of Culture.

References

Aristopoulos, A.: Register limits in singers; doct. diss. Venizelos Medical University (1983).

Coffin, B.: The sounds of singing (Pruett, Boulder 1976).

Cornut, G.; Lafon, J. C.: Etude acoustique comparative des phonèmes vocaliques de la voix parlée et chantée. Folia phoniat. *12:* 188–196 (1960).

Delattre. P.: Les modes du français. French Rev. *27* (1953).

Dutoit-Marco, M. L.: La voix (Favre, Paris 1985).

Germain, A.: Séassau, H.: Influence de la fréquence sur l'intelligibilité et les indices acoustiques des consonnes du français en voix chantée; DEA Diss. University Provence, Aix-en-Provence, pp. 1–115 (1982).

Gottfried, T. L.; Chew, S. L.: Intelligibilitiy of vowels sung by a countertenor. J. acoust. Soc. Am. *79:* 124–130 (1986).

Hoc, J. M.: L'analyse planifiée des données en psychologie. PUF, Coll. Le Psychologue, No. 91 (Paris, 1983).

Howie, J.; Delattre, P.: An experimental study of the effect of pitch on the intelligibility of vowels. NATS Bull. *18:* 6–9 (1962).

Husson, R.: Comment se forment les voyelles. La Nature, Paris *3267:* 249–257 (1957).

Husson, R.: Problèmes acoustiques et physiologiques posés par la formation des voyelles chantées. J. Physiol., Paris *50:* 328–331 (1958).

Janson, T.: Identical sounds and variable perception. Actes Congr. Phonol., Vienne 1980, pp. 215–222.

Johansson, C.; Sundberg, J.; Wilbrand, H.: X-ray study of articulation and formant frequencies in two female singers. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 4, pp. 117–134 (1982).

Miller, G. A.; Nicely, P. E.: An analysis of perceptual confusions among some English consonants. J. acoust. Soc. Am. *27:* 338–352 (1955).

Scotto Di Carlo, N.: Etude acoustique et auditive des facteurs d'intelligibilité de la voix chantée. Proc. 7th Int. Cong. Phonet. Sci., pp. 1017–1023 (Mouton, The Hague 1972).

Scotto Di Carlo, N.: Pourquoi ne comprend-on pas les chanteurs d'opéra? La Recherche *89:* 495–497 (1978).

Scotto Di Carlo, N.: Les problèmes d'intelligibilité dans le chant. Panorama du Médecin *1257* (1981).

Smith, L. A.; Scott, B. L.: Increasing the intelligibility of sung vowels. J. acoust. Soc. Am. *67:* 1795–1797 (1980).

Strange, W.: Verbrugge, R.; Shankweiler, D.; Edman, T.: Consonant environment specifies vowel identity. J. acoust. Soc. Am. *60:* 213–224 (1976).

Sundberg, J.: Vibrato and vowel identification. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 2/3, pp. 49–60 (1975).

Tarneaud, J.: Traité pratique de phonologie et de phoniatrie (Maloine, Paris 1961).

Triplett, W.: Investigation concerning vowel sounds on high pitches. NATS Bull. *50:* 6–8 (1967).

Received: October 16, 1985
Accepted: May 26, 1986

Nicole Scotto Di Carlo,
33, boulevard d'Arras,
F-13004 Marseille (France)