

Las interrogativas totales en el habla susurrada¹

Miguel Martín Echarri

Universidad de Burgos
mmecharri@ubu.es ORCID: <https://orcid.org/0000-0001-8654-1366>

Recibido: 29/09/2021; Aceptado: 09/11/2021; Publicado en línea: 16/09/2022

Cómo citar este artículo / Citation: Miguel Martín Echarri (2021). Las interrogativas totales en el habla susurrada. *Loquens*, 8(1-2), e082. <https://doi.org/10.3989/loquens.2021.e082>.

RESUMEN: El perfil entonativo es clave en la identificación de las interrogativas totales. Consecuentemente, su ausencia en el habla susurrada obliga a buscar otros rasgos que transmitan esa información. En referencia a otras lenguas, se ha atribuido a f_1 y f_2 la información prosódica en ausencia de f_0 , ascendiendo donde correspondería un ascenso de f_0 ; también se ha relacionado con la intensidad y con la duración.

El español presenta la ventaja de la estabilidad tímbrica de sus vocales, por lo que hemos analizado los formantes en tres enunciados pronunciado cada uno por tres informantes tanto en fonación modal como en susurro, en modalidad enunciativa e interrogativa. Aunque los resultados sí muestran un predominio del ascenso final de los formantes en las interrogativas, las excepciones no impidieron que se reconozca la modalidad en las grabaciones, lo que sugiere la responsabilidad de la variable tensión glótica.

Palabras clave: susurro; modalidades entonativas; interrogativa total; entonación; español.

ABSTRACT: *Yes-no questions in whispered speech.*- Intonative profiles are key to the identification of yes-no questions in many languages. Therefore, its absence in whispered speech makes it necessary to look for other features transmitting that information. Regarding other languages, f_1 and f_2 have been pointed as responsible of prosodic information in absence of f_0 , through a rise in those points where a rise of pitch would take place. It has also been related to intensity and duration.

The quality of vowels in Spanish is very stable, so we have chosen to analyse formants in three utterances, each one pronounced by three informants in phonation as well as in whispered speech, and both in declarative and interrogative modalities. Although results do show a prevalence of the final rise of both formants in interrogative utterances, all exceptions could be identified as interrogative too. This suggests other features are involved, probably tension of the vocal cords.

Key Words: whisper; intonative modalities; yes-no questions; intonation; Spanish.

¹ Este artículo debe mucho a la ayuda y asesoramiento de Miguel Roca Flores

1. INTRODUCCIÓN

La frecuencia fundamental (f_0) es un rasgo de los sonidos sonoros dependiente de la fonación al que todas las lenguas recurren y que cumple diversas funciones en cada una de ellas, desde la de distinguir segmentos (en las lenguas tonales) hasta la expresión de contenidos paralingüísticos y extralingüísticos. Hay acuerdo en reconocer que f_0 es la clave que permite al oyente construir la entonación, entendida consecuentemente como la variable perceptiva relacionada con los ascensos y descensos de f_0 , si bien hay otros aspectos que participan en su percepción. En lenguas no tonales como el español la entonación tiene, aparte de otras funciones estructuradoras e integradoras, una función fonológica: la de distinguir las modalidades oracionales. Son determinados perfiles entonativos asociados a pasajes concretos de los enunciados los que pueden aclarar la posición del hablante respecto al contenido expresado, al menos en aquellos casos en que esa posición no queda clara por otros medios (sintácticos, pragmáticos, etc.). Una interrogativa total puede quedar establecida sin ambigüedad mediante un ascenso brusco en los segmentos posteriores al último acento del enunciado.

Sin embargo, los hablantes a veces recurren al habla susurrada por exigencias de la situación, lo que supone la renuncia a la vibración de las cuerdas vocales, responsable de f_0 , y no por eso pierden su capacidad de expresar esas mismas modalidades entonativas que en voz modal asociamos a la frecuencia fundamental o, en términos más generales, siguen siendo capaces de reconocer esa información que debería corresponder a f_0 (Giet, 1956; Jensen, 1958; Miller, 1961; Abramson, 1972; Panconcelli-Calzia, 1955). Como oyentes hispanohablantes, somos capaces de identificar el mismo modelo interrogativo en un enunciado en el que no se oye f_0 .

Ello implica la presencia de otros rasgos que aparecen asociados a la línea de la frecuencia fundamental y que son redundantes con ella, de manera que son capaces de transmitir la misma información cuando aquella falta, o bien de rasgos especialmente generados para esa situación, porque deben tomar el relevo en su ausencia. La primera de esas opciones se centra en la hipótesis del correlato secundario (Heeren, 2015), según la cual los rasgos que permiten identificar la entonación en el habla susurrada podrían estar presentes ya en el discurso modal de manera redundante, pero resultar necesarios cuando falta f_0 . La prueba de la existencia de estos rasgos redundantes es que ha sido posible eliminar de manera sintética f_0 en enunciados pronunciados en voz modal en que los oyentes siguen reconociendo correctamente la entonación (Denes, 1959; Abramson, 1972). La otra hipótesis, complementaria pero no contradictoria con la primera, es la del correlato compensatorio, según la cual, ante las dificultades para expresar cierta información sin contar con f_0 , los hablantes podrían modificar los rasgos del sonido en uno o varios sentidos. Es obvio que el hablante tiene que adaptar sus emisiones a las necesidades que percibe en el oyente, debidas entre otras cosas al ruido circundante, su capacidad auditiva, su conocimiento del idioma, etc. Así,

es posible que el hablante refuerce algunos rasgos o incluso los modifique radicalmente para generar la sensación de entonación que f_0 no puede dar en el habla susurrada. Heeren y Van Heuven (2014) ofrecen pruebas que validan la hipótesis del correlato secundario, pero consideran que la mayoría de los casos se explican mejor por la del correlato compensatorio.

En cualquier caso, los rasgos que se han propuesto como responsables de codificar en el habla susurrada la información que normalmente se debe a f_0 se dividen en dos grandes tipos: los que se incluyen en el espectro (fundamentalmente los dos primeros formantes, pero también otras características espectrales) y los que no (la duración y la intensidad de los segmentos).

La idea de que hay características espectrales que se modifican de manera paralela a f_0 o que pueden modificarse para sustituirla dirige de manera bastante directa a los rasgos más llamativos del espectro: los formantes. En la búsqueda de los correlatos acústicos de la frecuencia fundamental se vio que la configuración de los formantes se modificaba de manera paralela a la que habría correspondido a la línea de f_0 (Meyer-Eppler, 1957; Fónagy, 1969): concretamente, las alturas de esos formantes ascendían en aquellos pasajes que correspondían a ascensos previsibles de f_0 (Higashikawa, Nakai, Sakakura y Takahashi, 1996). Esto se comprobó también por medio de la síntesis: los ascensos en f_1 y f_2 se percibían e interpretaban de un modo más claro en la misma dirección señalada por los estudios anteriores (Higashikawa y Minifie, 1999).

Pero no solo los formantes se modifican en paralelo a la f_0 : en ausencia de cambios sistemáticos en los formantes, parece claro que los oyentes recurren a otras pistas en el espectro. Así, Kong y Zeng (2006) analizan la capacidad de reconocer los tonos del mandarín en el habla susurrada y la atribuyen a las pistas que se encuentran en el espectro. Heeren y Lorenzi (2014), mostraron que era posible eliminar f_2 y f_3 de enunciados susurrados sin que la percepción por parte de los informantes se modificara, lo que hace pensar que hay otras características perceptibles que van en la misma dirección. En concreto, ese estudio apuntaba a la importancia del centro de gravedad (más alto en los puntos previsiblemente elevados de la entonación) y de la pendiente espectral (menos marcada en esos casos).

Respecto a los rasgos no espectrales, el que más atención ha logrado es la duración, que parece relacionarse directamente con la altura de f_0 : en mandarín susurrado, la buena identificación del tercer tono puede deberse a su duración relativamente más larga (Liu y Samuel, 2004). En este sentido, una mayor duración sería el sustituto de una mayor altura (Xiaonan, 1993), como se ha propuesto para el francés (Heeren y Lorenzi, 2014) o el holandés (Heeren y Van Heuven, 2014). Una explicación posible para el aumento de la duración es que sea un reflejo del alargamiento que sufren las vocales del español cuando existe una inflexión tonal importante en esa sílaba, lo que supondría que es un rasgo redundante de las modificaciones que exigen los otros rasgos que expresan esa información normalmente debida a la entonación: un correlato

secundario (la duración) del correlato compensatorio (las modificaciones en el espectro, probablemente).

Por último, la intensidad puede considerarse relacionada en el mismo sentido: los picos entonativos ausentes podrían verse sustituidos por picos de intensidad (Meyer-Eppler, 1957). Como ocurría con la duración, puede ser que el hablante aumente el gasto energético para señalar al oyente que merece la pena atender a las modificaciones que está realizando y podrían pasar desapercibidas, al menos en el caso de los perfiles más marcados (en español, el de las interrogativas totales, o el de las focalizaciones).

En resumen, parece bastante claro que los autores que han estudiado este problema en las más diversas lenguas tonales y no tonales optan prioritariamente por la idea de que f_1 y f_2 son el principal correlato compensatorio, puesto que son los que conllevan la principal información prosódica en ausencia de f_0 cuando los hablantes hacen ascender la altura de estos formantes en aquellos pasajes a los que corresponde un ascenso de la entonación. Pero los otros rasgos espectrales (el centro de gravedad y la pendiente espectral) también parecen ser solidarios con las variables que señalan la entonación, por lo que permiten su identificación en ausencia de f_0 . Por último, la implicación de la duración y la intensidad parece menor pero real.

Si las variaciones en las alturas de f_1 y f_2 son la clave, como defiende gran parte de la bibliografía, parece que será más fácil analizar enunciados en una lengua en la que los timbres vocálicos presenten poca variación en función de los acentos: frente al análisis de lenguas en las que la variación del timbre vocálico dificulta precisamente el estudio de su variabilidad en relación con esta función marginal del timbre como sustituto de la frecuencia fundamental, el español proporciona vocales muy estables en lo que concierne al timbre, así que cualquier variación puede ser atribuida a esta intención. En consecuencia, será fácil comparar en el espectrograma las alturas relativas de los formantes de dos o más apariciones del mismo fonema vocálico en un único enunciado. Esto permite introducir en el experimento enunciados algo más largos y con sentido lingüístico, frente a lo que es habitual en los estudios de la bibliografía, que se limitan a sílabas o incluso a sonidos aislados (Heeren, 2014, p. 132).

2. OBJETIVOS E HIPÓTESIS

La idea que hemos intentado verificar es que en el habla susurrada en español es posible encontrar recursos que sustituyan al menos los patrones entonativos más evidentes y necesarios para la comprensión de las modalidades. En este sentido, había que comprobar si los oyentes son efectivamente capaces de identificar la modalidad oracional tanto en habla susurrada como en voz modal, y también si lo hacen con dudas o con claridad.

A partir de ahí, hemos tratado de validar la hipótesis de que existe una correlación entre los valores de f_1 y f_2 , la duración silábica y la intensidad y el hipotético contorno de f_0 que le correspondería según la modalidad oracional: los valores aumentarían en paralelo a los aumentos previ-

sibles para f_0 , es decir, al final en los enunciados interrogativos, descendiendo en los declarativos.

Podemos detallar esta hipótesis: por un lado, queremos analizar el comportamiento del timbre vocálico, que previsiblemente se abrirá más en aquellas vocales a las que correspondería una frecuencia mayor de f_0 : las frecuencias de f_1 y f_2 serían paralelas a la que sería previsible para f_0 en una pronunciación con voz modal. En los enunciados interrogativos totales, f_1 y f_2 deberían ser claramente más graves en la última vocal tónica de cada enunciado que en la vocal postónica. Por el contrario, en los enunciados declarativos (o enunciativos) ambos formantes deberían descender hasta sus mínimos en la vocal postónica final.

Por su parte, la duración parece ser un factor relacionado también con la percepción de la entonación, pero cabe preguntarse si no se trata más bien de un aumento del tiempo requerido para pronunciar las inflexiones, es decir, los cambios de dirección de f_0 . Es lo que esperamos que ocurra en español: la modalidad interrogativa se caracteriza no tanto por la elevación del tono como por el cambio marcado desde la frecuencia más baja del enunciado hasta la más alta, y ese cambio requiere un mínimo de duración que se mantiene en el habla susurrada y que representa una pista muy clara para el oyente. Será interesante ver si en las interrogativas totales la duración de la última sílaba tónica es realmente menor que la de la sílaba final (previsiblemente la más aguda), o si, al contrario, la más larga es la tónica (previsiblemente la más grave), que es la que presenta la inflexión más clara.

Por último, hemos intentado también validar la hipótesis que relaciona las modalidades entonativas con las variaciones de intensidad. En los enunciados interrogativos totales, la fricción glótica que caracteriza el susurro podría ser más intensa en el final de la vocal postónica que en la última vocal tónica.

A partir de la bibliografía, parecía probable que estas variables aparecieran combinadas, de manera que pudiera concluirse que la identificación de la modalidad entonativa depende más del conjunto de los factores que de uno solo, aunque parecía posible que alguno de ellos fuera prioritario. En todo caso, si a un oyente le resulta posible identificar la modalidad entonativa en los enunciados susurrados grabados para nuestro experimento, deberemos suponer que hay algún rasgo del sonido que permite esa identificación.

3. METODOLOGÍA

Para verificar las hipótesis, se ha llevado un test de identificación a partir de unas grabaciones, y se han analizado esas grabaciones.

3.1. Estudio de percepción

Tres informantes (2 varones y 1 mujer, entre 35 y 70 años) han grabado una serie de enunciados que consistieron en la lectura de tres oraciones en cuatro versiones distintas cada una: a) voz modal, enunciativa; b) voz modal, interrogativa; c) susurro, enunciativa; d) susurro, in-

terrogativa. Cada una de las tres oraciones estaba pensada de tal manera que apareciese siempre el mismo fonema vocálico, que tuviera 7 sílabas y que terminase en una palabra llana (para facilitar el posterior análisis). Las oraciones fueron estas: “canta para la sala”, “este es el que perece” y “solo como los cocos”. Se excluyó la posibilidad de intentar componer oraciones con las otras dos vocales, /i/ y /u/, porque las combinaciones de palabras que pueden presentar exclusivamente estos fonemas habrían resultado demasiado extrañas para los informantes.

A los informantes se les dieron las instrucciones necesarias para que produjeran los enunciados sin explicarles en qué consistiría el análisis ni cuáles eran los objetivos de la investigación. El resultado podía ser más redundante que en casos tomados del habla real, pero eso era ante todo una ventaja: en la realidad del habla, las modalidades entonativas pueden ser redundantes con información que se encuentra en la situación, por lo que los hablantes pueden prescindir del esfuerzo que supone su realización. Además, analizar fragmentos de habla real habría supuesto la práctica imposibilidad de comparar los resultados: la probabilidad de encontrar un mismo enunciado en cuatro pronunciaciones distintas es mínima.

Las grabaciones se realizaron en una sola sesión, en una sala habilitada al efecto y con el micrófono incorporado a un ordenador portátil HP 250 G5. Las producciones constituyen una serie de 36 grabaciones que oscilan entre 1,26 y 3,83 segundos: 18 de ellas corresponden a enunciados susurrados y las otras 18 están dichas en voz modal; 18 son enunciativas y las otras 18 interrogativas; cada informante aparece en 12 de ellas; cada una de las tres vocales aparece en 12 de ellas. Las grabaciones no fueron manipuladas para modificar las alturas formánticas, las duraciones ni las intensidades ni ningún otro parámetro, de modo que aparecen las mismas características que se detallan más adelante, en el análisis de las producciones. Solo se modificó el orden de presentación para que fuera aleatorio (pero el mismo en las 10 escuchas).

A continuación se pidió a 10 jueces (5 varones y 5 mujeres, entre 35 y 75 años, hablantes de español norteamericano y sin problemas conocidos de audición) que escucharan todas las producciones de los tres informantes, para lo cual se ofreció a los informantes una hoja de respuestas en la que tenían que consignar una respuesta binaria para cada ítem. Escucharon la grabación y se les preguntó cuáles de los enunciados de la grabación eran interrogativos y cuáles enunciativos. También se les propuso que añadiesen una marca en los casos que les parecían dudosos. Así, puede considerarse que los posibles resultados son cuatro: acierto, acierto con dudas, error con dudas y error.

3.2. Estudio de producción

Para validar la segunda hipótesis se han analizado las grabaciones por medio de PRAAT (Boersma y Weenink, 2021), centrándonos en los parámetros que en las hipótesis hemos considerado relacionados con la identificación de las modalidades entonativas en los enunciados susurrados y en los realizados con voz modal.

Como punto de partida, se han realizado espectrogramas a partir de las grabaciones realizadas por los informantes. A continuación, esas grabaciones se han analizado con PRAAT, para centrar nuestra atención en la altura formántica de f_1 y f_2 (en hercios), la intensidad (en decibelios) y la duración (en segundos).

Para ello se ha partido de una segmentación cuidadosa que permitiera el análisis de los diferentes parámetros de nuestro interés. En esta investigación, la segmentación se ha realizado a partir fundamentalmente de la imagen espectrográfica recogida en PRAAT, pero se ha tenido cuidado de revisar uno por uno los segmentos delimitados, escuchando cuidadosamente el audio correspondiente y modificando la posición del límite (*boundary*) hasta ajustar en la medida de lo posible las características sonoras y los objetos visuales (fricciones desordenadas, barras de explosión, etc.) correspondientes al segmento. También se ha intentado asegurar que las características específicas de cada uno de los parámetros (formantes, intensidades, presencia o ausencia de f_0) coincidieran en sus límites. En las interrogativas se ha analizado también una octava unidad, que corresponde a la segunda mitad de la vocal postónica, en la idea de que el contraste será más evidente. En todo caso, este conjunto de decisiones se ha realizado de manera manual.

Partiendo de esa segmentación, se han analizado los parámetros que hemos considerado de interés para la validación de nuestras hipótesis.

En primer lugar, se ha analizado la relación entre la altura de la f_0 que encontramos en los enunciados pronunciados con voz modal y las alturas de f_1 y f_2 que aparecen en los susurrados. También se tiene en cuenta la comparación de los dos primeros formantes en los enunciados pronunciados en voz modal para comprobar si los cambios de timbre se dan también en esos casos.

El análisis de las alturas de los formantes se ha llevado a cabo valorando sobre todo los datos ofrecidos por el programa a partir de la delimitación de cada enunciado en los segmentos correspondientes, pero contemplando la posibilidad de modificar levemente la configuración de las variables a partir de las propuestas estandarizadas del programa (*settings* y *advanced settings*) en los casos en que los datos obtenidos estuvieran en contradicción con la imagen espectrográfica o con el sonido de la grabación. Así, hemos atendido en última instancia a la propia imagen recogida en el espectrograma como garante del análisis.

También se ha intentado verificar la hipótesis relativa a la tensión glótica: aunque no se llegue a la tensión necesaria para que se produzca una vibración en las cuerdas vocales, es posible que exista una diferencia de tensión en las cuerdas vocales que podría ser equivalente a la altura del sonido y que el oído lingüístico sea capaz de identificar. Partiendo de la idea de que esa tensión glótica va acompañada de un aumento de la intensidad, hemos complementado el estudio atendiendo a la intensidad comparada en los enunciados afirmativos e interrogativos y en los pronunciados con voz modal.

En cuanto a las duraciones, obviamente el principal problema estaba en las dificultades de la propia segmenta-

ción, que ya hemos visto que tiene importantes repercusiones en cualquiera de los otros parámetros analizados. Sin embargo, una vez tomada la decisión sobre la posición que correspondía a los límites entre los segmentos de cada uno de los enunciados grabados, el análisis de las duraciones ha resultado muy sencillo: PRAAT ofrece la posibilidad de seleccionar el segmento completo y señala su duración.

3.3. Estudio estadístico

Los datos se han definido como números totales. El estudio de dependencia entre los valores de diversos parámetros en las dos últimas vocales de cada enunciado se ha realizado mediante un análisis de correlaciones lineales; el análisis de homogeneidad de la independencia del ascenso de V6 a V7 frente a las voces modal o susurrada en interrogativas se ha llevado a cabo mediante el coeficiente de correlación de Pearson. Para la realización de los cálculos se ha usado el programa estadístico Statistical Package for the Social Sciences 15 (SPSS 15) (SPSS® Inc., Chicago, IL, USA).

4. RESULTADOS

Debemos atender primero a los resultados de la encuesta realizada para garantizar que los enunciados se identificaban adecuadamente con la modalidad prevista por sus locutores; después tendremos en cuenta los datos obtenidos a partir del análisis de las grabaciones de acuerdo con los distintos parámetros que pueden ayudarnos a validar las hipótesis.

4.1. El estudio de percepción

En la encuesta planteada para verificar que los enunciados grabados pueden ser identificados cómodamente por los oyentes, independientemente de estar pronunciados con voz modal o en susurro, el 100 % de los 18 enunciados pronunciados en voz modal fue identificado correctamente por los 10 informantes, que discriminaron perfectamente y sin dudas las modalidades enunciativa e interrogativa. En cambio, a la hora de discriminar los enunciados susurrados los resultados fueron un poco peores (Tabla 1): de los 180 datos, 175 fueron aciertos (97,2 %), frente a 5 fallos (2,8 %); y dentro de los aciertos, en 17 casos los informantes reconocieron haber tenido dudas al decantarse por una opción (9,4 %), mientras que el 87,8 % de los casos fueron identificados correctamente y sin dudas. De los 5 fallos, 3 consistieron en la identificación de una enunciativa cuando la pretensión del hablante era realizar una interrogativa, y los otros 2 en la identificación de una interrogativa en lugar de una enunciativa. En lo fundamental, estos datos apoyan la idea de que efectivamente los hablantes de español interpretan correctamente aunque no con total certeza la modalidad entonativa de los enunciados susurrados, y en concreto sirven para avalar los resultados obtenidos en el análisis de las grabaciones.

Si los analizamos por separado, podemos ver que (sumando las pronunciaciones de los tres locutores) los

Tabla 1: Resultados de las encuestas realizadas a 10 sujetos (S1-S10). En la primera columna aparece el orden (escogido aleatoriamente) en que se reprodujeron las grabaciones; en la segunda, la vocal presente en el enunciado; en la tercera, el informante (1, 2 o 3); en la cuarta, la modalidad del enunciado (“af” por “afirmativa”, “in” por “interrogativa”); en las siguientes, las respuestas en que los sujetos atribuían una modalidad a la grabación escuchada. Los signos de interrogación señalan las respuestas en que los sujetos reconocieron dudar en la respuesta.

Orden	Enunciado	Informante	Modalidad	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
1	/o/	1	in	in	in?	in	in	in	in	in	af	in	in
2	/o/	2	in	in	in?	af	in?	in	in?	in?	in?	in	in
3	/a/	2	af	af	af	in	af						
4	/e/	1	in	in	In	in	in	in	in	in	in	in	in
5	/a/	3	af	af	af	af	af	af	af	af	af	af	af
6	/e/	2	af	af	af	af	af	af	af	af	af	af	af?
7	/e/	3	in	in	in?	in	in	in	in	in	in	in	in
8	/e/	2	in	in	In	in	in	in	in	in	in?	in	in
9	/a/	3	in	in	in	in	in	in	in	in	in	in	in
10	/e/	1	af	af	in	af	af	af	af	af	af	af	af
11	/e/	3	af	af	af?	af	af	af	af	af	af	af	af
12	/o/	1	af	af	af	af	af	af	af	af	af	af	af
13	/o/	3	in	in	in	in	in?	in?	in?	in	in	in?	in?
14	/a/	2	in	in	in	in	af	in	in	in	in	in	in
15	/a/	1	af	af	af	af	af	af	af	af	af	af	af
16	/o/	3	af	af	af	af	af	af	af	af	af	af	af
17	/a/	1	in	in	in	in	in	in	in	in	in	in	in
18	/o/	2	af	af	af	af	af	af?	af	af	af?	af	af

enunciados con /a/ y /o/ provocaron más errores (2), pero /a/ no suscitó dudas, mientras que /o/ originó 13. El enunciado con /e/ dio lugar a un error y a cuatro casos dudosos. Esto podría sugerir que el timbre de las vocales puede presentar rasgos que interfieran de manera diferenciada en la identificación de la modalidad en voz susurrada.

No se ha dado ningún caso de error con dudas: parece que cuando los informantes tenían dudas realizaban un mayor esfuerzo de identificación que les permitía llegar a la conclusión adecuada. También hay pocos enunciados que hayan acumulado dudas y errores: solo un caso de /o/ ha ocasionado un error y cinco dudas; otro ha dado lugar a cinco identificaciones con dudas pero a ningún error (también interrogativa y con “solo”).

Respecto a los informantes, encontramos ciertas diferencias entre los tres: las grabaciones de la informante nº 1, de sexo femenino dieron lugar a 1 duda y 2 fallos; las del nº 2 (varón) provocaron 9 dudas y 3 fallos; las del nº 3 (varón) generaron 7 dudas.

Por último, es posible realizar el análisis de los datos a partir de la propia modalidad de los enunciados: encontramos 3 errores en la percepción de los enunciados

interrogativos, y 2 en la de los enunciativos. Es mayor el desequilibrio en la percepción de los enunciados dudosos: solo hay 4 dudas en los enunciados declarativos, frente a las 13 de los interrogativos. Tal vez la modalidad interrogativa consigue un grado de redundancia menor que la enunciativa, o la enunciativa es la menos marcada y es necesario conseguir una nitidez total de los rasgos interrogativos para que el oyente lo interprete sin dudas. En todo caso, se percibe en muchas de las grabaciones interrogativas susurradas una exageración de los rasgos en ausencia de una situación comunicativa que pudiera facilitar la interpretación del mensaje.

4.2. El estudio de producción

Empezaremos por señalar que en los enunciados pronunciados con voz modal todas las interrogativas fueron pronunciadas con un marcado ascenso en la sílaba final a partir de un mínimo en la última sílaba tónica (así que la relación $V7/V6^2$ da siempre resultados superiores a 1), mientras que las enunciativas mostraban una declinación más o menos previsible y una inflexión descendente más o menos marcada (de modo que la relación $V7/V6$ da siempre resultados inferiores a 1), aunque muchas veces la última sílaba resulta tan suave que la entonación no es perceptible. Es obvio que las condiciones del experimento (alejadas del habla real y de la comunicación en situación) forzaron pronunciaciones extremadamente obvias de esa modalidad, pero esa observación puede ser suficiente para indicar cuáles son los rasgos acústicos que los informantes relacionan con la entonación interrogativa y enunciativa.

4.2.1. Los dos primeros formantes

Los datos recogidos en la Tabla 2 muestran que las sílabas susurradas presentaron en la mayoría de los casos alturas mayores para f_1 que las sílabas equivalentes pronunciadas en voz modal (solo en un caso el valor de una vocal susurrada fue inferior a los de la misma vocal en voz modal: la segunda sílaba de la interrogativa con /a/, que no se recoge en la tabla pero sí en la Figura 4); pero la tendencia no es tan clara para f_2 , donde los valores se cruzan bastante: en el enunciado centrado en /o/ los datos son claros, pero no tanto en los otros enunciados. Estos resultados siguen la línea descrita por otros autores, como Kallail y Emanuel (1984). La Tabla 2 recoge los valores obtenidos para los dos primeros formantes de las dos últimas vocales en las grabaciones enunciativas.

En la comparación de los movimientos ascendentes o descendentes de los dos primeros formantes en las sílabas finales (la última tónica y la subsiguiente postónica), encontramos que en las enunciativas f_1 da resultados relativamente similares (predominantemente ascendentes) en las pronunciaciones con voz modal y las susurradas (Figuras 1, 2 y 3).

² En adelante, nos referiremos a la última vocal tónica como “V6”, y a la postónica como “V7”.

Tabla 2: Alturas de los dos primeros formantes de las vocales del totema de los enunciados declarativos, en hercios. Cada fila corresponde a una grabación, pero se subdivide en dos, una para cada formante; a su vez, cada columna ofrece información sobre el informante (1, 2 y 3), el fonema vocálico que aparece en cada enunciado (/a/, /e/, /o/), la voz (“Su.” para “susurrada” y “Mo.” para “modal”), el formante (f_1 y f_2) y las frecuencias correspondientes a V6 y V7, en Hz.

Informante	Sexo	Enunciado	Voz	Formante	V6	V7
1	M	/a/	Mo.	f_2	1521	1406
				f_1	838	749
1	M	/a/	Su.	f_2	1283	1499
				f_1	959	947
2	V	/a/	Mo.	f_2	1408	1208
				f_1	707	737
2	V	/a/	Su.	f_2	1660	1609
				f_1	1010	995
3	V	/a/	Mo.	f_2	1046	997
				f_1	625	584
3	V	/a/	Su.	f_2	1556	1332
				f_1	878	902
1	M	/e/	Mo.	f_2	2021	
				f_1	291	
1	M	/e/	Su.	f_2	2340	2377
				f_1	799	990
2	V	/e/	Mo.	f_2	1911	1893
				f_1	477	781
2	V	/e/	Su.	f_2	1826	1922
				f_1	770	935
3	V	/e/	Mo.	f_2	2035	1918
				f_1	443	580
3	V	/e/	Su.	f_2	1999	
				f_1	747	
1	M	/o/	Mo.	f_2	1009	1085
				f_1	482	419
1	M	/o/	Su.	f_2	1149	1325
				f_1	866	903
2	V	/o/	Mo.	f_2	1234	817
				f_1	542	547
2	V	/o/	Su.	f_2	1603	1688
				f_1	958	991
3	V	/o/	Mo.	f_2	792	953
				f_1	459	514
3	V	/o/	Su.	f_2	1828	1810
				f_1	933	904

Figura 1: Gráfico que resume las alturas en Hz de f_1 y f_2 en las pronunciaciones enunciativas de la oración “canta para la sala”.

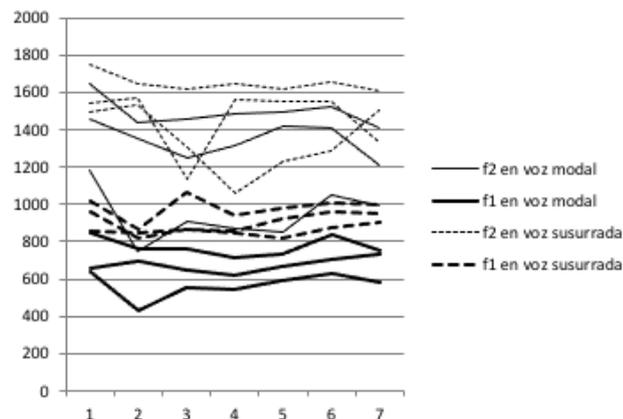


Figura 2: Gráfico que resume las alturas en Hz de f_1 y f_2 en las pronunciaciones enunciativas de la oración “este es el que perece”.

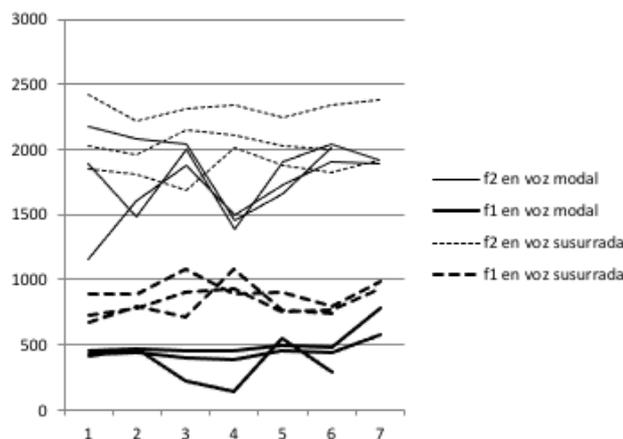
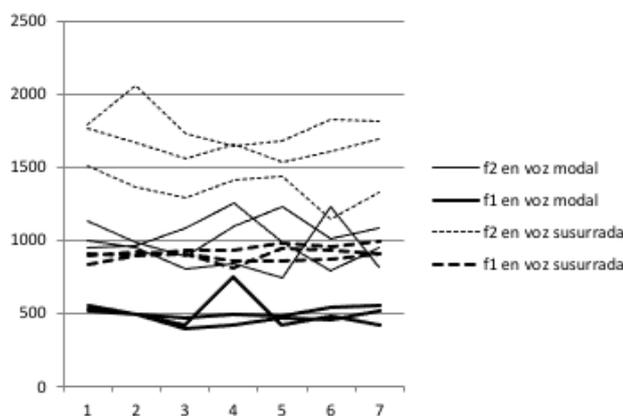


Figura 3: Gráfico que resume las alturas en Hz de f_1 y f_2 en las pronunciaciones enunciativas de la oración “solo como los cocos”.



Recogemos en la Tabla 3 un resumen de los ascensos y descensos en los finales que encontramos en las enunciativas.

Tabla 3: Suma de los finales (ascendentes, descendentes y dudosos) en las enunciativas.

		Ascendente	Descendente	Dudoso
Susurradas	f_2	5	3	1
	f_1	6	2	1
Modales	f_2	2	6	1
	f_1	5	3	1

Como puede verse, en las pronunciaciones enunciativas en voz modal se encuentran en f_1 5 ascensos frente a 3 descensos y 1 caso inaudible, mientras que en las susurradas vemos 6 ascensos, 2 descensos y 1 caso inaudible. En cambio, f_2 es predominantemente ascendente en los casos susurrados (5 ascensos, 3 descensos y 1 caso inaudible) pero descendente en los casos de fonación modal (2 ascensos, 6 descensos y 1 caso inaudible). Encontramos entonces en las pronunciaciones susurradas más finales ascendentes de lo que sería de esperar si suponemos que la altura de los formantes refleja la línea que habría seguido la frecuencia fundamental en caso de haberse pronunciado con voz modal.

Respecto a las interrogativas, la Tabla 4 recoge los datos extraídos de las grabaciones.

Tabla 4: Alturas de los dos primeros formantes de las dos vocales finales (V6 y V7) de los enunciados interrogativos totales, en hercios. Cada fila corresponde a una grabación, pero se subdivide en dos, una para cada formante, a partir de la quinta columna; a su vez, cada columna ofrece información sobre el informante (1, 2 y 3), el fonema vocálico que aparece en cada enunciado (/a/, /e/ y /o/), el sexo del informante (“V” para “varón”, “M” para mujer), la voz (“Su.” para “susurrada” y “Mo.” para “modal”), el formante (f_1 y f_2) y las frecuencias correspondientes a V6 y V7 y a la segunda mitad de la última vocal, en Hz.

Informante	Sexo	Enunciado	Voz	Formante	V6	V7	2ª mitad de V7
1	M	/a/	Mo.	f_2	1543	1441	983
				f_1	793	726	485
1	M	/a/	Su.	f_2	1209	1294	1625
				f_1	896	1002	1045
2	V	/a/	Mo.	f_2	1819	2099	1223
				f_1	773	604	470

Informante	Sexo	Enunciado	Voz	Formante	V6	V7	2ª mitad de V7
2	V	/a/	Su.	f_2	1523	1430	1602
				f_1	868	923	926
3	V	/a/	Mo.	f_2	798	1057	1038
				f_1	607	574	500
3	V	/a/	Su.	f_2	1558	1438	1407
				f_1	818	749	798
1	M	/e/	Mo.	f_2	2254	2128	1835
				f_1	440	467	348
1	M	/e/	Su.	f_2	2404	2555	2563
				f_1	862	818	790
2	V	/e/	Mo.	f_2	1829	1992	1504
				f_1	484	476	471
2	V	/e/	Su.	f_2	1875	1954	1934
				f_1	733	883	1040
3	V	/e/	Mo.	f_2	1720	1390	1565
				f_1	471	364	273
3	V	/e/	Su.	f_2	1942	2056	2151
				f_1	800	804	784
1	M	/o/	Mo.	f_2	1407	929	901
				f_1	597	521	488
1	M	/o/	Su.	f_2	1112	1303	1423
				f_1	870	847	874
2	V	/o/	Mo.	f_2	1162	837	865
				f_1	550	462	438
2	V	/o/	Su.	f_2	1754	1508	1480
				f_1	995	1003	1068
3	V	/o/	Mo.	f_2	872	913	966
				f_1	499	462	467
3	V	/o/	Su.	f_2	1205	1144	1216
				f_1	867	850	830

Estos datos pueden presentarse bajo la forma más fácilmente comprensible de tres gráficos, uno para los enunciados pronunciados con cada una de las vocales (Figuras 4, 5 y 6). Como puede verse, f_1 sigue una tendencia diferenciada, mayormente ascendente en las susurradas, casi siempre descendente en las pronunciadas con voz modal, mientras que f_2 sigue la misma tendencia, más veces ascendente en las susurradas, más veces descendente en las pronunciadas con voz modal.

Figura 4: Gráfico que resume las alturas en Hz de f_1 y f_2 en las pronunciaciones interrogativas de la oración “canta para la sala”.

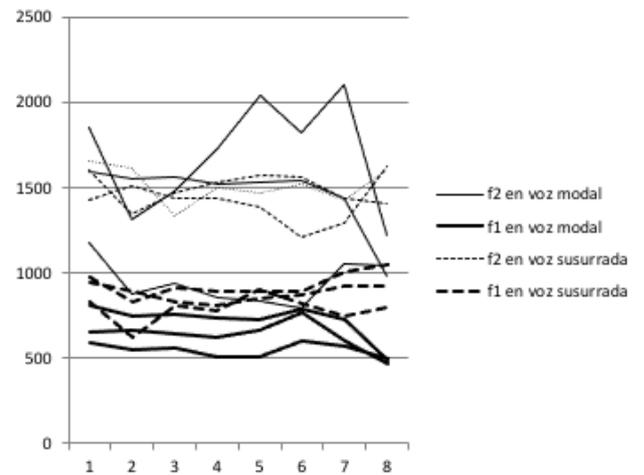


Figura 5: Gráfico que resume las alturas en Hz de f_1 y f_2 en las pronunciaciones interrogativas de la oración “este es el que parece”.

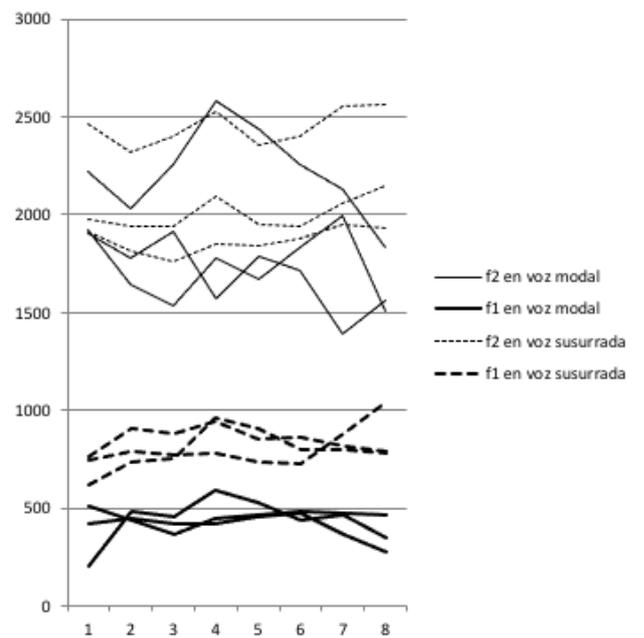
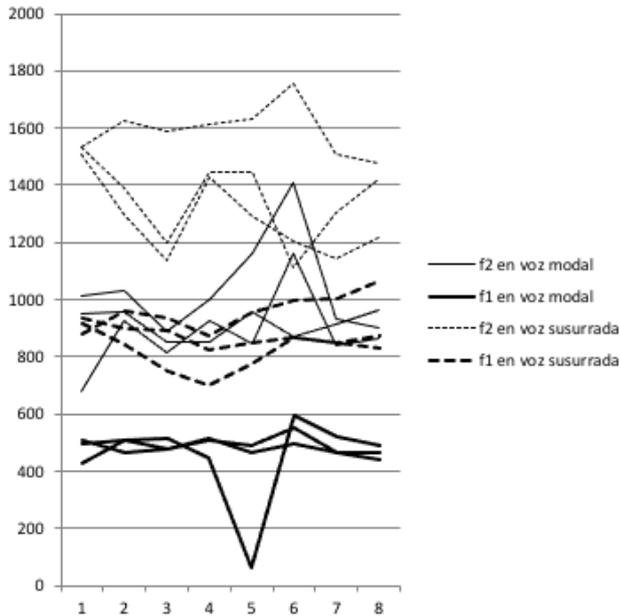
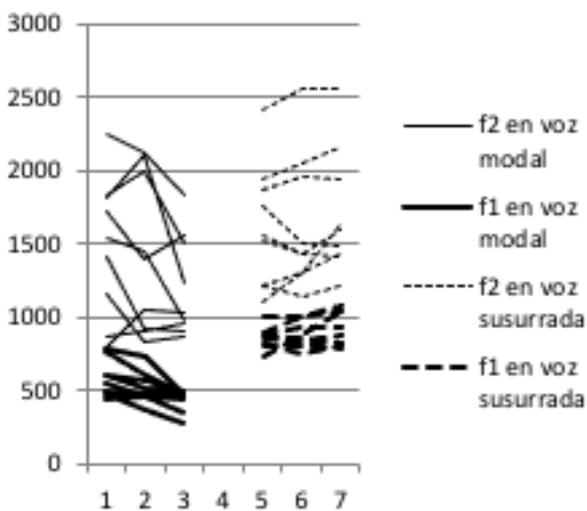


Figura 6: Gráfico que resume las alturas en Hz de f_1 y f_2 en las pronunciaciones interrogativas de la oración “este es el que perece”.



Podemos ofrecer otro gráfico (Figura 7) que recoge exclusivamente los finales comparados de las interrogativas modales y susurradas. En él se observa con claridad la mayor frecuencia de los formantes en voz susurrada, y también la tendencia ascendente que presentan en sus tonemas, frente a la descendente que caracteriza a los enunciados en voz modal. Pero también se ven excepciones a la tendencia, así como casos en los que se produce un cambio de dirección: descenso en el paso de la tónica a la postónica, seguido de un ascenso en la parte final de la postónica, o viceversa. Estos cambios de dirección son muy frecuentes en f_2 en voz modal (5 casos), pero también se dan en voz susurrada (3 casos para cada formante).

Figura 7: Gráfico que resume las alturas en Hz de f_1 y f_2 en V6, V7 y la segunda mitad de la vocal final en las pronunciaciones interrogativas.



Como hicimos con las enunciativas, ofrecemos a continuación un resumen de los finales ascendentes, descendentes y dudosos en las interrogativas (relación entre las dos últimas sílabas), añadiendo los mismos resultados en la relación entre la sílaba tónica y la segunda mitad de la postónica (Tabla 5).

Tabla 5: Suma de los finales (ascendentes, descendentes y dudosos) en las interrogativas, considerando f_1 y f_2 en la última vocal tónica (V6) frente a la postónica (V7); y en la última vocal tónica frente a la segunda mitad de la postónica.

		De V6 a V7			De V6 a 2ª mitad de V7		
		Ascendente	Descendente	Dudoso	Ascendente	Descendente	Dudoso
Susurradas	f_2	5	4	0	7	2	0
	f_1	5	4	0	5	4	0
Modales	f_2	4	5	0	2	7	0
	f_1	1	8	0	0	9	0

En estas pronunciaciones no se dio ningún caso de sílaba final inaudible (como sí ocurría con algunas de las enunciativas), lo que seguramente sugiere que en los enunciados interrogativos totales esos valores finales eran significativos para los informantes. Por otro lado, vemos que f_1 es siempre descendente en las modales (contra la dirección de f_0), pero es ascendente en 5 de las 9 pronunciaciones susurradas. Se trata de más de la mitad de los casos, pero está muy lejos de ser un rasgo generalizado. Respecto a f_2 , es también predominantemente descendente en las modales (7 de 9 casos) y, al contrario, es predominantemente ascendente en las susurradas (7 de 9). También aquí vemos una tendencia incluso más clara del formante a imitar la línea que habría correspondido a f_0 , pero sin que la línea de ese final sea sistemática.

4.2.2. Las intensidades

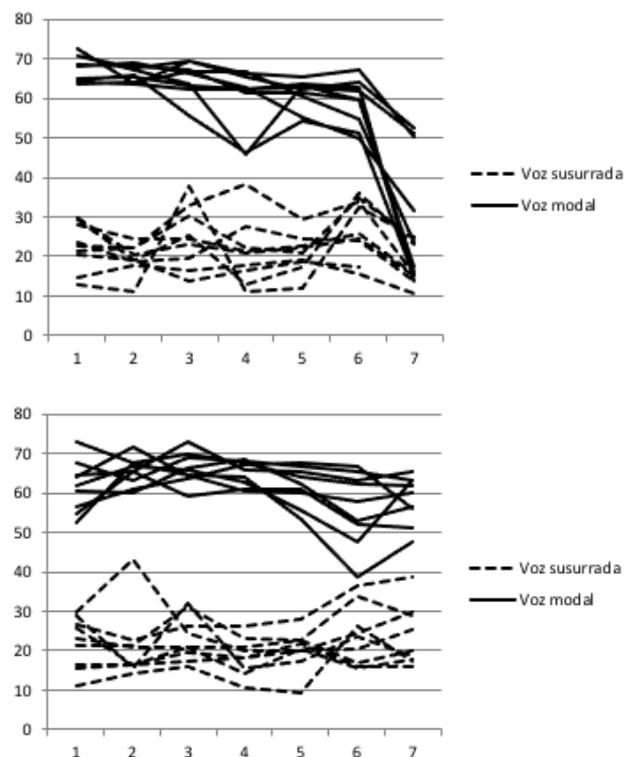
En relación con la comparación de las intensidades, recogemos los resultados en la Tabla 6. Es fácil comparar las diferencias entre los enunciados susurrados y los pronunciados en voz modal, de una intensidad llamativamente mayor; pero, aparte de esto, unos y otros parecen compartir en alto grado las diferencias de intensidad entre las dos últimas vocales: en los dos casos, las enunciativas se relacionan con una decidida caída de la intensidad en la vocal final, y en los dos casos las interrogativas suponen un relativo mantenimiento de esa intensidad, con 4 enunciados en que la intensidad desciende frente a otros 5 en que asciende, pero en general con diferencias de intensidad poco pronunciadas.

Tabla 6: Intensidades medias de las 2 vocales finales de los enunciados, en decibelios. Cada fila corresponde a una grabación; cada columna ofrece información sobre el informante (1, 2 y 3), el fonema vocálico que aparece en cada enunciado (/a/, /e/ y /o/), la modalidad (“En.” para “enunciativa” e “In.” para “interrogativa”), la voz (“Su.” para “susurrada” y “Mo.” para “modal”) y la intensidad media en V6 y V7, en dB. Por último, incluimos la relación entre V7 y V6.

Informante	Sexo	Enunciado	Modalidad	Voz	V6	V7	V7/V6
1	M	/a/	En.	Mo.	66,98	50,09	0,75
2	V	/a/	En.	Mo.	49,81	31,71	0,64
3	V	/a/	En.	Mo.	63,83	52,61	0,82
1	M	/e/	En.	Mo.	62,69	13,82	0,22
2	V	/e/	En.	Mo.	51,22	15	0,29
3	V	/e/	En.	Mo.	59,46	16	0,27
1	M	/o/	En.	Mo.	61,79	50,98	0,83
2	V	/o/	En.	Mo.	54,51	23,44	0,43
3	V	/o/	En.	Mo.	59,6	17,22	0,29
1	M	/a/	In.	Mo.	65,43	63	0,96
2	V	/a/	In.	Mo.	38,49	47,39	1,23
3	V	/a/	In.	Mo.	62,37	62,02	0,99
1	M	/e/	In.	Mo.	66,79	56,17	0,84
2	V	/e/	In.	Mo.	52,11	51,16	0,98
3	V	/e/	In.	Mo.	58,01	59,95	1,03
1	M	/o/	In.	Mo.	52,88	56,31	1,06
2	V	/o/	In.	Mo.	47,51	63,27	1,33
3	V	/o/	In.	Mo.	63,25	65,43	1,03
1	M	/a/	En.	Su.	33,57	15,5	0,46
2	V	/a/	En.	Su.	24,51	15,29	0,62
3	V	/a/	En.	Su.	25,61	15,63	0,61
1	M	/e/	En.	Su.	23,95	13,98	0,58
2	V	/e/	En.	Su.	15,6	10,83	0,70
3	V	/e/	En.	Su.	17,13		0
1	M	/o/	En.	Su.	34,79	24,43	0,70
2	V	/o/	En.	Su.	32,44	24,93	0,77
3	V	/o/	En.	Su.	36,06	23,11	0,64
1	M	/a/	In.	Su.	36,45	38,64	1,06
2	V	/a/	In.	Su.	20,22	25,45	1,26
3	V	/a/	In.	Su.	24,37	29,98	1,23
1	M	/e/	In.	Su.	16,86	20,02	1,19
2	V	/e/	In.	Su.	15,78	15,78	1
3	V	/e/	In.	Su.	15,61	17,7	1,13
1	M	/o/	In.	Su.	34	28,76	0,85
2	V	/o/	In.	Su.	26,16	17,13	0,66
3	V	/o/	In.	Su.	23,53	18,88	0,80

Por otro lado, como muestran los gráficos siguientes, resulta claro que las diferencias de intensidad en los finales son más abultadas en los enunciados pronunciados con voz modal. Podemos comprobarlo en los enunciados declarativos e interrogativos (Figura 8).

Figura 8: Gráficos que recogen las intensidades de las 7 vocales de las enunciativas (gráfico superior) y las interrogativas totales (inferior).



4.2.3. Las duraciones

Respecto a las duraciones, en la Tabla 7 se muestran los resultados para las vocales finales, y en la Figura 9 los correspondientes a todas las sílabas de todos los enunciados.

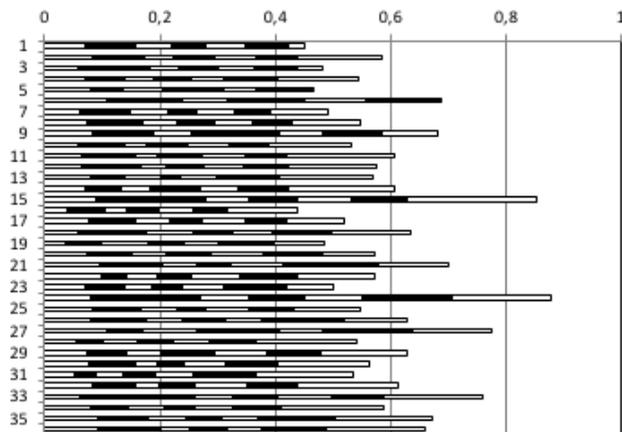
Tabla 7: Duraciones de V6 y V7 y relación entre las dos. Cada fila corresponde a una grabación; cada columna ofrece información sobre el informante (1, 2 y 3), su sexo (“V” para “varón” o “M” para “mujer”), el fonema vocálico que aparece en cada enunciado (/a/ para “canta”, /e/ para “este” y /o/ para “solo”), la modalidad (“En.” para “enunciativa” e “In.” para “interrogativa”), la voz (“Su.” para “susurrada” y “Mo.” para “modal”). A continuación, la duración de V6 y V7 en segundos y la relación entre ellas.

Informante	Sexo	Enunciado	Modalidad	Voz	V6	V7	V7/V6
1	M	/a/	En.	Mo.	0,074	0,029	0,389
2	V	/a/	En.	Mo.	0,072	0,147	2,040
3	V	/a/	En.	Mo.	0,075	0,044	0,584
1	M	/e/	En.	Mo.	0,094	0,141	1,495
2	V	/e/	En.	Mo.	0,098		0,000
3	V	/e/	En.	Mo.	0,131		0,000
1	M	/o/	En.	Mo.	0,064	0,1	1,555
2	V	/o/	En.	Mo.	0,068	0,121	1,775
3	V	/o/	En.	Mo.	0,102	0,098	0,961
1	M	/a/	In.	Mo.	0,068	0,145	2,149
2	V	/a/	In.	Mo.	0,071	0,186	2,613
3	V	/a/	In.	Mo.	0,079	0,151	1,920
1	M	/e/	In.	Mo.	0,108	0,164	1,522
2	V	/e/	In.	Mo.	0,086	0,185	2,149
3	V	/e/	In.	Mo.	0,098	0,225	2,305
1	M	/o/	In.	Mo.	0,057	0,122	2,164
2	V	/o/	In.	Mo.	0,07	0,1	1,423
3	V	/o/	In.	Mo.	0,101	0,139	1,370
1	M	/a/	En.	Su.	0,099	0,088	0,892
2	V	/a/	En.	Su.	0,102	0,091	0,896
3	V	/a/	En.	Su.	0,165	0,121	0,729
1	M	/e/	En.	Su.	0,1	0,133	1,324
2	V	/e/	En.	Su.	0,109	0,081	0,745
3	V	/e/	En.	Su.	0,157	0,17	1,084
1	M	/o/	En.	Su.	0,079	0,114	1,440
2	V	/o/	En.	Su.	0,142	0,109	0,769

Informante	Sexo	Enunciado	Modalidad	Voz	V6	V7	V7/V6
3	V	/o/	En.	Su.	0,156	0,137	0,883
1	M	/a/	In.	Su.	0,081	0,173	2,133
2	V	/a/	In.	Su.	0,095	0,148	1,561
3	V	/a/	In.	Su.	0,09	0,159	1,774
1	M	/e/	In.	Su.	0,109	0,168	1,548
2	V	/e/	In.	Su.	0,087	0,174	2,000
3	V	/e/	In.	Su.	0,092	0,171	1,855
1	M	/o/	In.	Su.	0,087	0,177	2,024
2	V	/o/	In.	Su.	0,133	0,169	1,273
3	V	/o/	In.	Su.	0,112	0,17	1,511

A continuación, el conjunto de los datos en un gráfico (Figura 9).

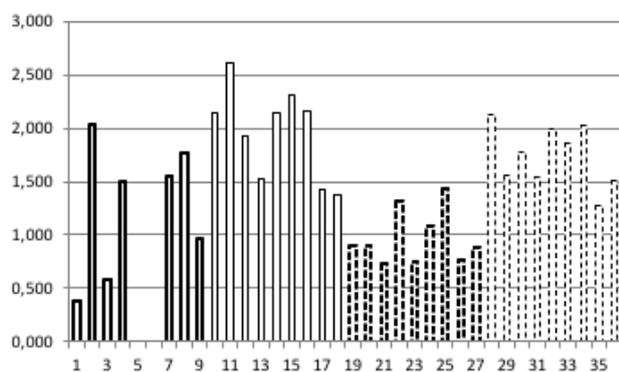
Figura 9: Gráfico que recoge las duraciones de las 7 vocales de los enunciados analizados. Cada sección de una barra, en blanco y negro de manera alterna, corresponde a una vocal. Los enunciados están ordenados: modales enunciativos (1-9), modales interrogativos (10-18), susurrados enunciativos (19-27) y susurrados interrogativos (28-36).



Los resultados son coherentes con lo previsto: la vocal final átona suele ser más larga que la media de las vocales del enunciado (aunque hay dos casos en voz modal en los que no se percibe bien, porque también suelen ser menos intensas), pero en el caso de las interrogativas esa duración es marcadamente mayor, superando siempre con claridad la media. En las interrogativas susurradas esa tendencia supera a las pronunciadas con voz modal.

Se atendió también a la relación entre la duración de V7 y V6 (Figura 10), para verificar la posibilidad de que el alargamiento de la vocal postónica se realice por contraste con la tónica y, sobre todo, para comprobar si la sílaba con entonación más aguda en voz modal es siempre la más larga en voz susurrada, tanto en las enunciativas (en las que corresponde a la última sílaba tónica) como en las interrogativas totales (en las que corresponde a la sílaba final, postónica). Y encontramos en los resultados que en las susurradas interrogativas todas las vocales finales superan con claridad la duración de las vocales tónicas (con valores que van desde 1,27 hasta 2,13), en las susurradas enunciativas los valores son generalmente inferiores, pero no sistemáticos y no siempre menores de 1 (van desde 0,73 hasta 1,44). Es decir, aunque en las enunciativas susurradas la postónica es también a menudo más larga que la última tónica, es clara la tendencia a alargar la postónica en las interrogativas.

Figura 10: Gráfico que recoge la relación entre las duraciones de V7/V6. Las columnas lisas corresponden a los enunciados pronunciados en voz modal; las discontinuas a los susurrados; las gruesas a las enunciativas; las finas a las interrogativas.



Por otro lado, los resultados obtenidos en los enunciados en voz modal van en la misma dirección: la relación entre las duraciones de la vocal final y la última tónica también da valores mayores en las interrogativas que en las enunciativas (lo que en principio hace pensar que se trata de un correlato secundario), pero con más entrecruzamientos que en las susurradas, y con un valor de correlación mucho más bajo. También encontramos en este caso los dos enunciados en que la sílaba final resultó inaudible, con lo que fue imposible establecer los límites de la vocal.

Para nuestros informantes, parece claro que se requiere en las interrogativas una cierta duración que permita realizar la inflexión que define el tonema y que distingue esta modalidad de la enunciativa; eso es coherente con la circunstancia de que en algunos casos la sílaba final se encuentre tan debilitada que resulta inaudible, porque en los enunciados modales la duración de la vocal final se relacionaría con la ausencia de un segmento posterior que obligue a un corte drástico, y no con el alargamiento necesario para señalar la inflexión.

Pero se sigue de lo anterior que es necesario relacionar la duración con la intensidad: si multiplicamos la relación que se da entre V7 y V6 en intensidades y en duraciones, obtenemos los resultados que recogemos en la Tabla 8.

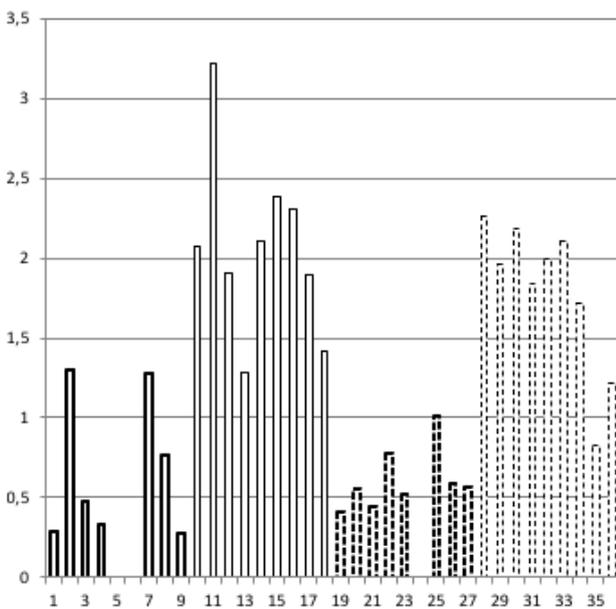
Tabla 8: Valores correspondientes al producto entre duraciones e intensidades (tomando en estas la proporción entre la vocal final, V7, y la penúltima, V6).

Informante	Sexo	Enunciado	Modalidad	Voz	Duración (V7/V6) x intensidad (V7/V6)
1	M	/a/	En.	Mo.	0,29
2	V	/a/	En.	Mo.	1,3
3	V	/a/	En.	Mo.	0,48
1	M	/e/	En.	Mo.	0,33
2	V	/e/	En.	Mo.	0
3	V	/e/	En.	Mo.	0
1	M	/o/	En.	Mo.	1,28
2	V	/o/	En.	Mo.	0,76
3	V	/o/	En.	Mo.	0,28
1	M	/a/	In.	Mo.	2,07
2	V	/a/	In.	Mo.	3,22
3	V	/a/	In.	Mo.	1,91
1	M	/e/	In.	Mo.	1,28
2	V	/e/	In.	Mo.	2,11
3	V	/e/	In.	Mo.	2,38
1	M	/o/	In.	Mo.	2,3
2	V	/o/	In.	Mo.	1,89
3	V	/o/	In.	Mo.	1,42
1	M	/a/	En.	Su.	0,41
2	V	/a/	En.	Su.	0,56
3	V	/a/	En.	Su.	0,45
1	M	/e/	En.	Su.	0,77
2	V	/e/	En.	Su.	0,52
3	V	/e/	En.	Su.	0
1	M	/o/	En.	Su.	1,01
2	V	/o/	En.	Su.	0,59
3	V	/o/	En.	Su.	0,57
1	M	/a/	In.	Su.	2,26
2	V	/a/	In.	Su.	1,96
3	V	/a/	In.	Su.	2,18

Informante	Sexo	Enunciado	Modalidad	Voz	Duración (V7/V6) x intensidad (V7/V6)
1	M	/e/	In.	Su.	1,84
2	V	/e/	In.	Su.	2
3	V	/e/	In.	Su.	2,1
1	M	/o/	In.	Su.	1,71
2	V	/o/	In.	Su.	0,83
3	V	/o/	In.	Su.	1,21

En la Figura 11 puede notarse la llamativa diferencia entre las columnas correspondientes a los enunciados declarativos e interrogativos; pero también puede notarse que los pocos casos en que los valores de las interrogativas susurradas descienden hasta cruzarse con los de las enunciativas (los dos últimos), son precisamente los que han causado más dudas en el estudio de percepción. Sin embargo, como se verá en el análisis estadístico, la correlación de esta variable con la f_0 de las interrogativas en voz modal es débil.

Figura 11: Gráfico que recoge el producto entre V7/V6 (dura-ciones) y V7/V6 (intensidades). Las columnas lisas corresponden a los enunciados pronunciados en voz modal; las discontinuas a los susurrados; las gruesas a las enunciativas; las finas a las interrogativas.



4.2.4. Análisis estadístico

A partir de los datos recogidos hasta aquí, se ha intentado encontrar posibles correlaciones lineales entre los distintos valores que podrían suponer un correlato secundario o compensatorio. Para ello, se han realizado análisis de correlaciones lineales entre los valores de las alturas de f_1 y f_2 en las dos últimas vocales de cada enunciado.

En primer lugar, para los enunciados enunciativos en voz modal, las correlaciones lineales entre V7 y V6, entre V7-V6 y V6, y entre [(V7-V6)/V7] y V6, en f_1 y f_2 por separado, correlaciones de entre las cuales la única que tiene un coeficiente de correlación alto es V7/V6 para f_2 : $r = 0,92$, lo que indica una correlación lineal directa fuerte entre V6 y V7 para f_2 . La relación lineal sería: $V7 = 1284,63 + 0,88 (V6 - 1369,50)$. En segundo lugar, las mismas posibilidades en los enunciados enunciativos en voz susurrada, correlaciones de entre las cuales aparecen tres que tienen un coeficiente de correlación alto: V7//V6 para f_2 , correlación directa $r = 0,92$; (V7-V6)//V6 para f_1 , correlación inversa $r = -0,88$; y ((V7-V6)/V7)/V6 para f_1 , correlación inversa $r = -0,88$. Las ecuaciones que las relacionan son: $V7 = 1695,25 + 0,988 (V6 - 1655,63)$ // $(V7 - V6) = 49,25 - 0,87 (V6 - 896,63)$ // $((V7 - V6)/V7) = 5,13 - 0,09 (V6 - 896,63)$.

A continuación, para los enunciados interrogativos en voz modal, se analizaron las correlaciones entre V7 y V6, entre V7-V6 y V6, y entre [(2ª mitad de V7)-V6] y V6, en f_1 y f_2 por separado, de entre las cuales resultaron coeficientes de correlación alto para V7//V6 para f_1 y f_2 , correlación directa $r = 0,9$ y $r = 0,86$; y (2ª mitad de V7 - V6) // V6, correlación inversa $r = -0,8$. Las relaciones lineales serían: en f_2 $V7 = 1420,67 + 0,95(V6 - 1489,33)$ // en f_1 $V7 = 517,33 - 0,73 (V6 - 579,33)$ // en f_1 . $D2 = -141,56 - 0,67 (V6 - 579,33)$. Respecto a la modalidad interrogativa en voz susurrada, de esas mismas correlaciones resultó tener un coeficiente de correlación alto solamente V7/V6 para f_2 : $r = 0,95$, que indica una correlación lineal directa fuerte entre V6 y V7 para f_2 . La relación lineal sería $V7 = 1631,33 + 1,20 (V6 - 1620,22)$.

Más interés podría tener el análisis de las correlaciones entre V6 susurrada y V6 modal, y entre V7 susurrada y V7 modal, en f_1 y f_2 por separado, pero todas esas relaciones tienen coeficientes de correlación no próximos a 1. Por lo tanto, no existe correlación lineal entre V6(M) y V6(S) ni entre V7(M) y V7(S) tanto para f_1 como para f_2 . A su vez, de las correlaciones entre [V7 - V6] susurradas y [V7 - V6] modales, en f_1 y f_2 por separado, la relativa a f_1 tiene un coeficiente aceptable que permite encontrar una relación lineal directa, moderadamente fuerte, entre (V7(M)-V6(M)) // (V7(S) - V6(S)) en f_1 . La relación sería: $V7(S) - V6(S) = 29 + 0,39 ((V7(M) - V6(M)) - 28,71)$. Respecto a las correlaciones entre V6 susurrada y V6 modal, y entre V7 susurrada y V7 modal, y entre [(V7-V6)/V7] y V6, en f_1 y f_2 por separado, ninguna de ellas tiene un coeficiente de relación próximo a 1, por lo que no puede encontrarse correlación lineal entre V6(M) y V6(S) ni entre V7(M) y V7(S) tanto para f_1 como para f_2 . Por último, de las

correlaciones entre [V7 - V6] susurradas y [V7 - V6] modales, en f_1 y f_2 por separado, ninguna tiene coeficiente de correlación cercano a 1.

Respecto a la variable intensidad, se analizaron las correlaciones lineales entre V6 (susurrada) y V6 (modal), entre V7 (susurrada) y V7 (modal), y entre [(V7-V6) susurrada] y [(V7-V6) modal], en primer lugar para las enunciativas, pero todos los coeficientes de correlación resultaron pequeños; y en segundo lugar para las interrogativas, en que aparecieron coeficientes de regresión de 0,73 para V6(S)/V6(M) y V7(S)/V7(M), una correlación lineal directa moderada entre ellos cuyas ecuaciones son: $V6(S) = 21,93 + 0,38 (V6(M) - 49,87)$ y $V7(S) = 21,63 + 0,40 (V7(m) - 51,64)$.

También se analizó la correlación entre algunos parámetros de las interrogativas susurradas y modales con la f_0 de las modales: esta es la relación que podría apuntar con más claridad a los correlatos más importantes para la identificación de las modalidades. En todos esos casos se tenía en cuenta el cociente V7/V6. Entre estos parámetros se incluyeron f_1 y f_2 en las susurradas enunciativas e interrogativas, pero de esas cuatro posibilidades tres resultaron prácticamente inexistentes: la correlación entre f_0 en las enunciativas modales y f_2 en las enunciativas susurradas ($r=0,006$); la correlación entre f_0 en las interrogativas modales y f_1 en las interrogativas susurradas ($r=0,217$); y la correlación entre f_0 en las interrogativas modales y f_2 en las interrogativas susurradas ($r=-0,149$). Solo resultó un coeficiente de correlación lineal inversa muy fuerte entre f_0 en las enunciativas modales y f_1 en las enunciativas susurradas ($r=-0,977$).

También se analizó la correlación entre f_0 en las modales con la duración, con la intensidad y con el producto de la duración y la intensidad (se trataba siempre de los cocientes V7/V6). En las enunciativas, las tres correlaciones resultaron negativas débiles ($r = -0,460$ para la duración; $-0,522$ para la intensidad; $-0,544$ para el producto de la duración y la intensidad); en las interrogativas, las tres resultaron inexistentes ($r = 0,097$ para la duración) o débiles ($r = 0,479$ para la intensidad; $r = 0,400$ para el producto de intensidad por duración).

Aparte de estos análisis, se llevó a cabo el contraste de Pearson de la independencia del ascenso de V6 a V7 frente a la voz modal o susurrada en interrogativa. Teniendo en cuenta que el valor de contraste es 6,63, y que todos los valores que relacionaban el ascenso de V6 a V7 en cada uno de los dos primeros formantes eran menores que 6,63 (en las enunciativas con las tres vocales y en las interrogativas con /a/ y con /o/), debemos aceptar, para un nivel de confianza del 99 % la hipótesis de que, para cualquiera de los dos primeros formantes en cualquiera de los enunciados (salvo los interrogativos con /e/, para los que no había datos suficientes), el ascenso de V6 a V7 es independiente del rasgo modal o susurrado.

También se analizó el contraste de Pearson de la independencia del número de aciertos en el test de percepción respecto del informante y de la modalidad: te-

niendo en cuenta que el valor de contraste es 10,6 y el valor estadístico 0,012, aceptamos, para un nivel de confianza del 99 %, la hipótesis de que hay independencia del informante con relación a la modalidad para el número de aciertos de los jueces. Del mismo modo, con un valor estadístico de 0,080 (también inferior al valor de contraste, 10,6) aceptamos, para un nivel de confianza del 99 % la hipótesis de que hay independencia del enunciado (centrado en el fonema /a/, /e/, /o/) con relación a la modalidad para el número de aciertos de los jueces.

En resumen, los datos disponibles no son suficientes para establecer con claridad correlaciones entre las variables analizadas. Pero nos interesa tener en cuenta que es posible encontrar una relación lineal directa, moderadamente fuerte, entre (V7 modal-V6 modal) y (V7 susurrada - V6 susurrada) en f_1 .

5. DISCUSIÓN

Los resultados obtenidos siguen en general las líneas descritas por los autores citados en la bibliografía: de acuerdo con todos los autores desde Giet (1956) hasta Heeren (2015), los jueces identificaron con pocas dudas los enunciados pronunciados como interrogativos y como enunciativos; los formantes son más agudos en general en los enunciados susurrados, tal como se ve desde Higashikawa, M., Nakai, K., Sakakura, A. y Takahashi, H. (1996); hay una tendencia a elevarse en los puntos asociados a elevaciones de f_0 , y esa tendencia es más marcada en f_1 que en f_2 (en consonancia con Heeren y Lorenzi, 2014); hay un aumento en la duración de las sílabas finales que es más marcado en las interrogativas totales que en las enunciativas, lo que apoya lo mostrado por Liu y Samuel (2004), Heeren y Lorenzi (2014) y Heeren y Van Heuven (2014); por último, hemos registrado picos de intensidad relacionados con los agudos (de acuerdo con los mismos autores).

Con las condiciones de nuestro estudio, y aunque los datos reunidos pueden ser insuficientes para sacar conclusiones fiables, hay que concluir que ninguno de los rasgos parece suficiente para expresar en solitario las modificaciones de la entonación, relacionadas en los enunciados en voz modal con f_0 : porque encontramos varias excepciones y solapamientos en todas las variables analizadas, que sin embargo no han impedido una buena identificación de los patrones por parte de los oyentes. Por el contrario, puede buscarse una combinación de características acústicas susceptible de identificarse con la entonación y de ofrecer al oyente una impresión equivalente a la frecuencia fundamental del sonido (esta es la opción más coherente con lo que se acepta generalmente en la bibliografía). Entre estas combinaciones posibles, en este estudio destaca la relación entre duración e intensidad, que puede explicarse como un énfasis en la duración por oposición al alargamiento neutro que caracteriza a las enunciativas, si bien la combinación de estas dos variables en el tonema no pudo relacionarse con las variaciones de f_0 en las interrogativas pronunciadas en voz modal.

6. CONCLUSIONES

La conclusión más importante de este estudio es que hemos podido ampliar su objeto de análisis a grabaciones de enunciados más complejos que los que son comunes en la bibliografía, hecho que ha sido posible debido en parte a la estabilidad formántica de las vocales del español. Eso ha acercado un poco nuestro estudio a la realidad del habla, siempre dentro de unos límites marcados por el propio objeto de estudio: ha sido necesario proponer la pronunciación de enunciados claramente cerrados y en condiciones de laboratorio, muy lejos de la improvisación y fuera de cualquier contexto comunicativo real, pero en todo caso bastante más cerca de las muestras de habla que en los casos de síntesis de vocales o de grabaciones de sílabas sin significado, que son los más comunes en estudios previos.

Los resultados no son concluyentes, pero redundan en la necesidad de relacionar distintos parámetros. Entre estas combinaciones, los casos de anomalías en el producto entre duración e intensidad pueden relacionarse con la dificultad de identificar correctamente la modalidad, aunque el estudio estadístico no encontró correlación con f_0 en las interrogativas modales analizadas.

7. REFERENCIAS

- Abramson, A. S. (1972). Tonal experiments with whispered Thai. En A. Valdman (Ed.), *Papers on Linguistics and Phonetics to the Memory of Pierre Delattre* (pp. 29–55). La Haya: Mouton.
- Boersma, P., y Weenink, D. (2021). PRAAT: Doing phonetics by computer [programa] (versión 6.1.53) <http://www.praat.org/>
- Denes, P. (1959). A preliminary investigation of certain aspects of intonation. *Language and Speech*, 2, 106–122. <https://doi-org.ubu-es.idm.oclc.org/10.1177/002383095900200204>
- Fónagy, J. (1969). Accent et intonation dans la parole chuchotée. *Phonetica* 20, 177–192.
- Giet, F. (1956). Kann man in einer Tonsprache fluestern? *Lingua* 5, 372–381. [https://doi-org.ubu-es.idm.oclc.org/10.1016/0024-3841\(55\)90029-2](https://doi-org.ubu-es.idm.oclc.org/10.1016/0024-3841(55)90029-2)
- Heeren, W. F. L. y Lorenzi, C. (2014). Perception of prosody in whispered French. *Journal of the Acoustical Society of America*, 135, 2026–2040. <https://doi-org.ubu-es.idm.oclc.org/10.1121/1.4868359>
- Heeren, W. F. L. (2014). Does boundary tone production in whispered speech depend on its bearer? Exploring a case of tonal crowding in whisper. En J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N. O. Schiller y E. van Zanten (Eds.), *Above and Beyond the Segments. Experimental Linguistics and Phonetics* (pp. 131–143). Amsterdam: John Benjamins Publishing Company.
- Heeren, W. F. L. y Van Heuven, V. J. (2014). The interaction of lexical and phrasal prosody in whispered speech. *Journal of the Acoustical Society of America*, 136(6), 3272–3289. <https://doi-org.ubu-es.idm.oclc.org/10.1121/1.4901705>
- Heeren, W. F. L. (2015). Vocalic correlates of pitch in whispered versus normal speech. *Acoustical Society of America*, 138 (6), 3800–3810. <https://doi.org/10.1121/1.4937762>
- Higashikawa, M. y Minifie, F. D. (1999). Acoustic-perceptual correlates of “whisper pitch” in synthetically generated vowels. *Journal of Speech, Language and Hearing Research*, 42(3), 583–591.
- Higashikawa, M., Nakai, K., Sakakura, A. y Takahashi, H. (1996). “Perceived Pitch of Whispered Vowels—Relationship with Formant Frequencies: A Preliminary Study”. *Journal of voice*, 10 (2), 155–158. <https://doi-org.ubu-es.idm.oclc.org/10.1044/jslhr.4203.583>
- Jensen, M. K. (1958). Recognition of Word Tones in Whispered Speech. *Word*, 14, 187–196. <https://www.tandfonline.com/doi/pdf/10.1080/00437956.1958.11659663>
- Kallail, J. J. y Emanuel, F. W. (1984). An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects. *Journal of Phonetics*, 12, 175–186.
- Kong, Y.-Y. y Zeng, F.-G. (2006). Temporal and spectral cues in Mandarin tone recognition. *Journal of the Acoustical Society of America*, 120(5), 2830–2840. <https://doi-org.ubu-es.idm.oclc.org/10.1121/1.2346009>
- Liu, S. y Samuel, A. G. (2004). Perception of Mandarin lexical tones when f_0 is neutralized. *Language and Speech*, 47, 109–138. <https://doi-org.ubu-es.idm.oclc.org/10.1177/00238309040470020101>
- Meyer-Eppler, W. (1957). Realization of prosodic features in whispered speech. *Journal of the Acoustical Society of America*, 29, 104–106. <https://doi-org.ubu-es.idm.oclc.org/10.1121/1.1908631>
- Miller, J. D. (1961). Word tone recognition in Vietnamese whispered speech. *Word*, 17, 11–15. <https://doi-org.ubu-es.idm.oclc.org/10.1080/00437956.1961.11659743>
- Panconcelli-Calzia, G. (1955). Das Flüstern in seiner physio-pathologischen und linguistischen Bedeutung. *Lingua*, 4, 369–378. [https://doi-org.ubu-es.idm.oclc.org/10.1016/0024-3841\(54\)90071-6](https://doi-org.ubu-es.idm.oclc.org/10.1016/0024-3841(54)90071-6)
- Xiaonan S. S. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech*, 36(4), 415–433. <https://doi.org/10.1177/002383099303600404>