



El Efecto McGurk en Consonantes Oclusivas sonoras y Sordas del Español de Chile

The McGurk Effect in Stop Voiced And Voiceless Consonants in Spanish From Chile

Silvana Andrea D'Ottone Campana¹

¹ Licenciada en Letras Pontificia Universidad Católica de Chile
Magíster en Lingüística Pontificia Universidad Católica de Chile
Docente Universidad Santo Tomás
sadotton@uc.cl

Artículo recibido: 28- Julio- 2011 **Aceptado:** 24- Octubre- 2011 **Publicado:** 22- Noviembre- 2011

RESUMEN

El efecto McGurk es un fenómeno perceptivo audiovisual de efecto fuerte, en el cual la información visual de una sílaba es incongruente con las pistas auditivas presentadas a los espectadores. Esto genera confusiones perceptivas en el espectador. El presente trabajo tiene como finalidad contribuir a los estudios sobre percepción de las consonantes oclusivas sonoras y sordas del español de Chile. Para ello, se trabaja dentro de un paradigma McGurk, mediante un experimento en el cual 84 participantes son expuestos a 8 estímulos audiovisuales incongruentes, con el fin de comprobar este fenómeno en hablantes chilenos y medir sus respuestas. A partir de ellas, se proponen algunas hipótesis sobre cuáles serían los rasgos determinantes para la percepción de las consonantes oclusivas presentadas. Además, se comparan los resultados obtenidos con los reportados para otras lenguas. Los participantes muestran uso tanto de las pistas visuales y auditivas al menos en la mitad de los casos; no obstante, el patrón de respuesta ante la serie sonora y la serie sorda es el inverso al reportado en otros estudios.

Palabras clave: Efecto McGurk, percepción audiovisual, consonantes oclusivas, estímulos incongruentes

ABSTRACT

The McGurk effect is a robust audiovisual phenomenon in which the visual information of a syllable does not match the auditory cues presented to the listeners. This generates perceptive confusions on the listeners. The former study's aim, is to contribute to the perception research of voiced and voiceless stop consonants in Chilean Spanish. Therefore, an experiment within a McGurk paradigm is performed, in which 84 participants are exposed, individually, to 8 incongruent audiovisual stimuli, in order to demonstrate whether this phenomenon could be proved among Chilean speakers and to assess their answers. Thus, some hypothesis are given to explain which the decisive features in the perception of the presented stop consonants could be. Furthermore, the results are compared to those obtained in other languages. Participants show the use of both visual and auditory cues. Nonetheless, the answer pattern for voiced and voiceless consonants is different from those reported in the consulted literature.

Keywords: McGurk effect, audiovisual perception, stop consonants, incongruent stimuli

1 Introducción

El efecto McGurk es una ilusión perceptiva audio-visual de efecto fuerte. En este, la información acústico-auditiva de un fonema es incongruente con la información articulatoria (video) presentada al espectador. Al intentar procesar ambas modalidades, la persona tiende a combinar o fusionar las informaciones de los dos estímulos. El ejemplo más común consiste en presentar a un espectador un video con la imagen de frente de la articulación de la sílaba [ga], pero con el audio sustituido por el de [ba]. El resultado debería ser que el sujeto perciba [da]; es decir, ninguno de los fonemas que están siendo presentados, sino uno nuevo. Los autores han utilizado el nombre de “fusión” para denominar a este fenómeno.

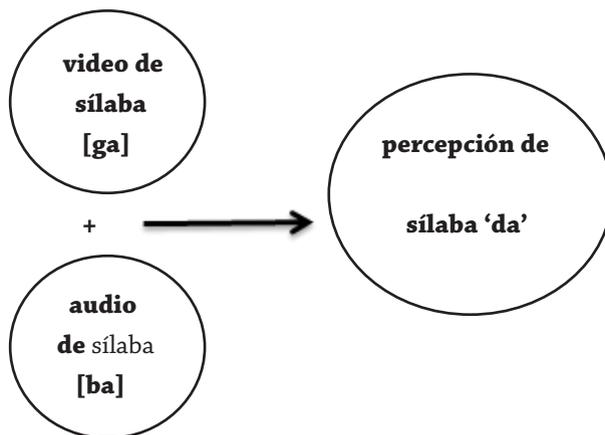


Fig. 1 Esquema representativo del efecto McGurk

Según la teoría en torno a este efecto, tanto en esta como en otras ilusiones, las respuestas de los oyentes-espectadores dependerían de cómo integran la información visual y auditiva que perciben simultáneamente. Es decir, las respuestas estarían determinadas por los rasgos acústicos y articulatorios (al menos por aquello percibido superficialmente) de las informaciones incongruentes y cómo estos son procesados por quien ve y escucha.

Dicho de otro modo, los rasgos acústicos, por un lado, y las pistas visuales de la articulación de las consonantes, por otro lado, formarían parte de la respuesta perceptiva de la persona expuesta al efecto McGurk. La pregunta es ¿Cuáles de dichos rasgos y pistas son determinantes para la percepción de un fonema?

Esta investigación nace como un intento de aportar a la teoría en relación al efecto McGurk en el español de Chile, dada la limitada cantidad de información existente. Además, en vista de las explicaciones que se han

tratado de hacer desde otros ámbitos de conocimiento (Baart y Vroomen, 2010; Boersma, 2006; Gentilucci y Cattaneo, 2005; Nicholls, Searle y Bradshaw, 2003; Omata y Mogi, 2008; Skipper, Wassenhove, Nusbaum y Small, 2007), parece necesaria la participación de la lingüística, ya que ella puede ofrecer una nueva perspectiva para comprender el fenómeno. Cabe señalar también la relevancia que tendría el obtener un mayor conocimiento del efecto McGurk, debido a los distintos ámbitos en los que puede ser aplicado (sicología, neurociencia, enseñanza de lenguas, entre otros). Esto justifica la necesidad de explorarlo de manera interdisciplinaria; en este caso, la perspectiva del estudio es principalmente fonética, pero se toman conocimientos provenientes de otras áreas.

A raíz de lo anterior, la primera inquietud que surge es cómo responden los hablantes del español de Chile ante la percepción de estímulos audio-visuales incongruentes. Del mismo modo, es interesante preguntarse por qué ante una realización incongruente (McGurk) los sujetos perciben, mayoritariamente, un fonema y no otro; es decir, qué marcas o rasgos acústicos y articulatorios conducen a la percepción de un fonema consonántico oclusivo determinado. Más aún, lo que resulta particularmente intrigante es qué rasgos acústicos o articulatorios permiten la percepción de un fonema que no ha sido presentado, ni en la modalidad auditiva, ni en la modalidad visual, como sucede en el caso de las llamadas fusiones. Para intentar resolver estos cuestionamientos, es preciso en primera instancia presentar algunos antecedentes relevantes para la investigación.

1.1 El Efecto McGurk: historia y aplicaciones

Como se ha señalado antes, el efecto McGurk es un fenómeno de efecto fuerte bajo información audiovisual incongruente, en el cual la información que se obtiene de la lectura de labios interfiere con la percepción auditiva del estímulo presentado. Esta mezcla de estímulos tiene como resultado que los espectadores tiendan a combinar o fusionar ambas informaciones, lo que se conoce como ilusión perceptiva. Este tipo de ilusión fue descubierta accidentalmente por Harry McGurk mientras revisaba videos doblados de las consonantes /ba/ y /ga/ en secuencias congruentes e incongruentes. El fenómeno se hizo conocido en un artículo de 1976 como “Efecto McGurk”. A partir de entonces, se incrementaron las discusiones e investigaciones en torno a la percepción audio-visual y la percepción del habla en general.

El fenómeno se ha estudiado principalmente desde la neurociencia y la neuropsicología. Son precisamente estas áreas las que han hecho un mayor uso de la evidencia existente en torno al fenómeno y de sus consecuencias en la percepción audio-visual, para tratar de resolver un asunto que ha sido

debatido durante mucho tiempo y que aún hoy no encuentra solución: cómo se produce la percepción del habla. En esta ocasión, no obstante, la perspectiva es lingüística, y en este sentido es preciso conocer lo que será el principal objeto de estudio del estudio: las consonantes oclusivas.

1.2 Consonantes oclusivas

Las consonantes oclusivas pueden definirse como sonidos consonánticos en los cuales los órganos de la articulación se contactan de manera que se obstruye momentáneamente la salida del aire por la boca (Gili Gaya, 1961). Este proceso articulatorio consta de tres fases. La primera corresponde a la fase implosiva, en que los órganos activos se ponen en movimiento para adoptar el gesto articulatorio de contacto correspondiente. En la segunda fase, oclusiva, se cierra el paso del aire puesto que los órganos articulatorios se encuentran unidos. Este cierre del paso del aire genera un aumento de la presión del mismo, el cual sale de forma abrupta en la tercera fase de explosión (Martínez y Fernández, 2007).

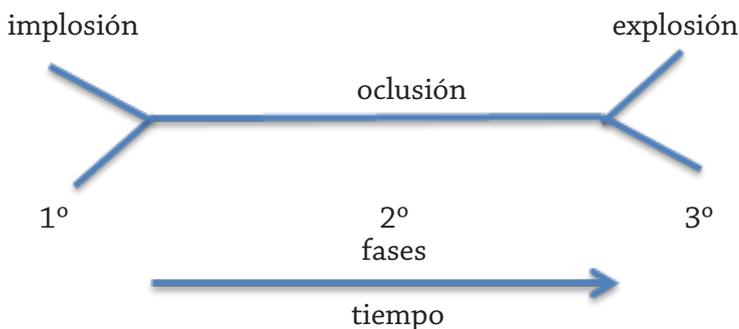


Fig. 2 Fases de una consonante oclusiva. Extraído de Martínez y Fernández (2007)

Según su punto de articulación, se consideran bilabiales aquellas en que hay un completo contacto de los labios como en p y b; las dentales, en cambio, se producen por el contacto del ápice de la lengua, como órgano activo, con la cara interna de los incisivos superiores, como es el caso de d y t. Las velares, se articulan mediante el contacto del postdorso de la lengua y el paladar, como sucede al pronunciar k o g (Gili Gaya, 1961).

En términos articulatorios, las consonantes oclusivas suelen ser clasificadas como sonoras o sordas, dependiendo de si existe o no vibración de las cuerdas vocales en el momento de su ejecución. En las oclusivas sordas, no se produce vibración de las cuerdas vocales, ni en el momento de implosión

ni en el de explosión. En las consonantes sonoras, la sonoridad o vibración de las cuerdas vocales comienza en el momento de implsión y no se interrumpe en las fases siguientes (cierre y explosión) (Gili Gaya, 1961). En cuanto a su descripción acústica, Ladefoged (1996) explica, en relación con las consonantes sonoras que, acústicamente, son los movimientos del segundo y tercer formante lo que caracteriza a estas consonantes.

En cuanto a la perceptibilidad de las oclusivas, Martínez y Fernández (2007) indican que esta involucra tanto lo audible de la fase explosiva del sonido, como lo que no es percibido (ausencia de sonido). El momento de cierre, según un estudio llevado a cabo por el autor, puede tener distinta duración y de ello dependerá que el oyente perciba la consonante como sonora, sorda o geminada. En opinión del autor, esta segunda fase de oclusión es necesaria y suficiente para la discriminación de un sonido.

En Chile actualmente existe escasa investigación en el área de la percepción del habla, al menos desde la lingüística. Los estudios existentes se han preocupado más de los componentes acústicos o articulatorios de las consonantes oclusivas. Esto puede deberse a la dificultad que supone generar pruebas de percepción que resulten eficaces y confiables para el objetivo propuesto. No obstante, el efecto McGurk puede ser una buena alternativa para los estudios de percepción del habla, tal como se verá a continuación.

1.3 Algunas investigaciones en torno al efecto McGurk

Dentro del paradigma McGurk, Colin *et al* (2002) investigan el rol de la sonoridad y de la intensidad de los estímulos auditivos presentados en la percepción del efecto McGurk. Los autores concluyen que el número de ilusiones aumenta en la medida en que disminuye la intensidad del sonido (de 70 a 40 dB). Además, señalan que el fenómeno cobra mayor fuerza cuando el estímulo visual es de buena calidad y las condiciones de audio están degradadas.

Otro dato interesante arrojado por esta investigación es que el número de fusiones es bastante menor al número de combinaciones, lo cual, plantean los autores, podría deberse a las propiedades fonéticas de la lengua en estudio. No obstante, al parecer esta asimetría numérica entre ambos tipos de ilusiones se debería al efecto del tipo de consonante: las consonantes sordas presentan un mayor número de combinaciones que las sonoras y, si bien no es estadísticamente significativo, el número de fusiones es un tanto mayor para las consonantes sonoras. Este mismo patrón es reportado por McDonald & McGurk (1978) y Sukiyaama & Tohkura (1991).

Entre otras variables que han sido estudiadas, Behne *et al* (2007) investigan el impacto de la edad en el fenómeno McGurk con el objetivo de determinar si existen diferencias entre adultos jóvenes (19-30 años) y adultos de mediana edad (49-60 años) en el uso de las pistas auditivas, visuales o audio-visuales en la identificación del lugar de articulación del estímulo percibido. Los resultados obtenidos permiten determinar que las diferencias entre los grupos etarios se deben a un distinto uso de las pistas visuales y no a la agudeza para percibir los estímulos.

En concordancia con la investigación de Colin *et al*, se puede observar también en este estudio, que las consonantes sonoras conducen a más fusiones que las consonantes sordas. En síntesis, esta investigación indicaría que el procesamiento de la información sensorial va cambiando con la edad y que el uso de la información visual se incrementa en la adultez, probablemente por la mayor experiencia adquirida con el tiempo.

Omata y Mogi (2008) también investigan el rol de la experiencia en la percepción audio-visual para explicar la magnitud del efecto McGurk en los espectadores. En este sentido, los autores explican que es posible que el efecto McGurk sea un sub-producto de la adquisición del lenguaje, impulsado por la exposición a estímulos audio-visuales coherentes. A través de modelos de reconocimiento del habla desarrollados en el área de la ingeniería, llegan a la conclusión de que el efecto McGurk sería una consecuencia del hecho de que la clasificación de fonemas es adquirida en el contexto de información audio-visual. Además, señalan que si bien el lenguaje se puede adquirir solo mediante la audición, la integración AV implicaría una ventaja en este proceso.

La importancia de esta integración audio-visual, señalan los autores, se relaciona con el proceso de la articulación del habla, posiblemente a través de la significatividad de la observación de la producción del habla, la cual activa el sistema espejo. Esto se relaciona también con el hecho de que el incremento en edad incrementa el efecto, como indican los resultados de Behne *et al*. (2007), lo cual además se relaciona con la experiencia articulatoria.

En base a lo anterior es posible constatar la utilidad del efecto McGurk para los estudios de percepción. Si bien aún presenta muchas incógnitas, trabajar dentro de un paradigma McGurk parece ser una buena posibilidad para contribuir a la descripción perceptiva de los fonemas oclusivos del español de Chile.

2. Metodología:

El experimento se lleva a cabo con 84 sujetos (18-25 años) hablantes nativos de español de Chile que no presentan ningún tipo de impedimento auditivo o visual (en este último caso, si se trata de un problema corregido, el sujeto es igualmente válido para el procedimiento).

El procedimiento consiste en la presentación de estímulos audio-visuales, incongruentes, los cuales muestran un hablante de género femenino (solo la parte de la boca y nariz) pronunciando sílabas que incluyen las consonantes oclusivas sonoras y sordas del español de Chile: /ba da ga/ /pa ta ka/. En dichos estímulos lo presentado visualmente no coincide con lo presentado acústicamente. La edición de los videos se realizó en el programa iMovie para Mac. El experimento constó de 8 estímulos:

SERIE SONORA:	SERIE SORDA:
baV daA	paV taA
baV gaA	paV kaA
daV baA	taV paA
gaV baA	kaV paA

Tabla i Estímulos sonoros y sordos presentados

La A mayúscula se sitúa junto a la consonante que es usada como estímulo de audio, mientras que la V mayúscula indica la consonante que corresponde a la articulación presentada en el video.

El experimento fue diseñado en el programa Psyscope para Mac. En este se presentan las instrucciones para realizar la tarea. Luego, los estímulos se presentan en modalidad de pantalla completa. A continuación de cada estímulo, se le despliegan las opciones de respuesta. El participante debe pulsar en el teclado la letra correspondiente a la alternativa que cree haber percibido. Una vez que se realiza una elección, el programa arroja el estímulo siguiente. Las respuestas de los sujetos son grabadas por el programa.

Para la realización del experimento, cada sujeto es situado frente a la pantalla de un computador Mac de 13 pulgadas. Además, se le entregan audífonos. De este modo, los participantes pueden ver y oír el estímulo simultáneamente. Luego de la aparición de cada estímulo, el sujeto debe elegir lo que cree haber percibido entre una serie de 7 opciones. A continuación, se presenta un ejemplo de las opciones dadas a los participantes.

En el caso de que el estímulo presente una sílaba 'ba' en la modalidad visual, junto con una 'da' auditiva:

1. ba
2. da
3. bda
4. dba
5. ga
6. pa
7. ta

Las opciones son ordenadas de manera aleatoria.

La aplicación del instrumento se realizó en 4 sesiones de aproximadamente una hora y media cada una en las aulas de una universidad, con alumnos de la misma institución, pertenecientes a diversas áreas de estudio.

Los datos obtenidos son analizados, en un primer momento, de forma descriptiva mediante la lectura de los gráficos elaborados a partir de los mismos. Para ello, los resultados se agrupan de tres formas: en primer lugar se analiza el porcentaje de cada tipo de respuesta para la totalidad de estímulos; en segundo lugar, el porcentaje de ilusiones en contraste con el de respuestas auditivas y otras respuestas (visuales con confusión de sonoridad) en función de cada tipo de estímulo; por último, el porcentaje de tipo de respuestas por cada estímulo. Asimismo, este análisis se realiza primero con la serie sonora y, luego con la serie sorda.

En un segundo momento del estudio, se realiza un análisis estadístico más fino mediante una prueba de chi-cuadrado, con el objetivo de observar si existe una relación de independencia entre las variables 'tipo de respuesta' y 'tipo de estímulo presentado'.

Para efectos de este análisis, se consideran como ilusiones las respuestas en que exista una combinación o fusión de ambos estímulos, o bien, aquellas en que el participante reporte la percepción del estímulo visual, sin que haya confusión de sonoridad. Las respuestas auditivas corresponden a la percepción del estímulo auditivo, sin considerar el estímulo visual. La opción "otro" corresponde a respuestas en que se considera el estímulo visual o el auditivo con confusión de sonoridad; es decir, el sujeto solo ha considerado la información del movimiento de los labios, pero no la información auditiva que debería informarle acerca de la sonoridad del estímulo.

3. Análisis de los datos

3.1 Análisis de datos para la serie de consonantes sonoras.

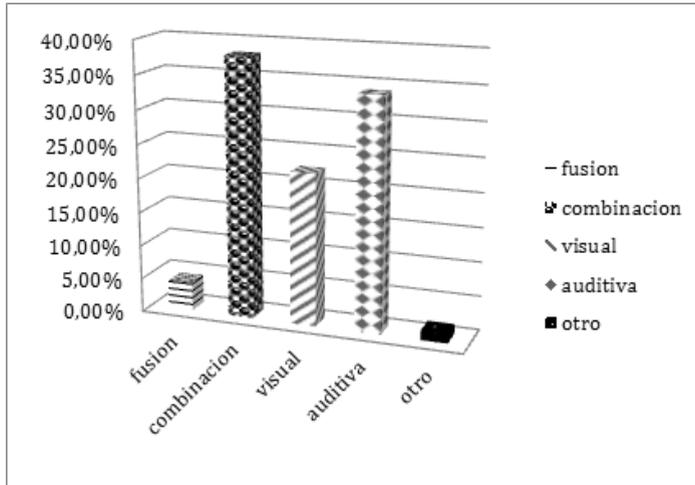


Gráfico 1 Porcentajes de los tipos de respuesta en la serie de consonantes sonoras

Los resultados obtenidos para las consonantes oclusivas sonoras, expresados en el gráfico 1, muestran que la cantidad de respuestas que implican algún tipo de ilusión, corresponde al 64,58% del total. Las respuestas que consideran solo el estímulo auditivo obtienen un porcentaje de 33,92%. Aquellas en que hay confusión de sonoridad alcanzan solo un 1,48%.

Como se puede observar, las respuestas de combinación alcanzan el porcentaje más alto de 38,09%. En cuanto a las respuestas de tipo visual sin confusión de sonoridad, estas representan el 22,32% del total. Las respuestas de fusión alcanzan el 4,16%.

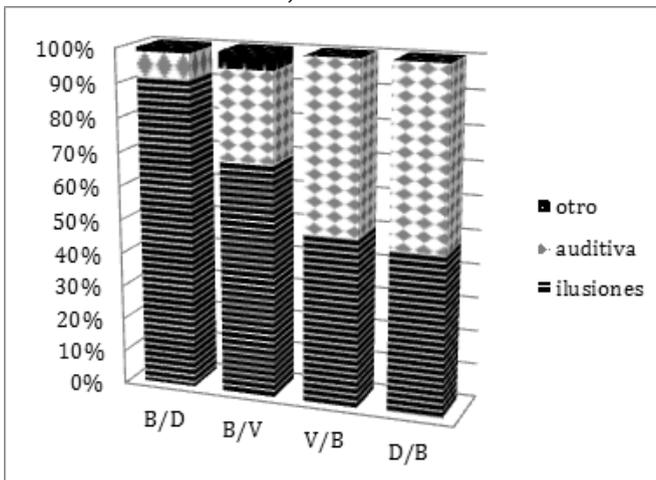


Gráfico 2 Porcentajes de ilusiones por tipo de estímulo presentado en la serie de consonantes sonoras

El gráfico 2 presenta los datos obtenidos respecto al porcentaje de ilusiones en consonantes sonoras para cada tipo de estímulo presentado. En él se observa que existe un mayor porcentaje de ilusiones en los estímulos con consonante bilabial como componente visual. En el estímulo B/D, constituido por la consonante bilabial como video y la consonante dental como audio, el porcentaje de ilusiones es de 91,66%, mientras que en B/V, en el cual el audio corresponde a una velar, es de 69,04%. En los casos en que la consonante bilabial ocupa el lugar del audio, el porcentaje de respuestas auditivas es similar al de ilusiones y fluctúa alrededor del 50%.

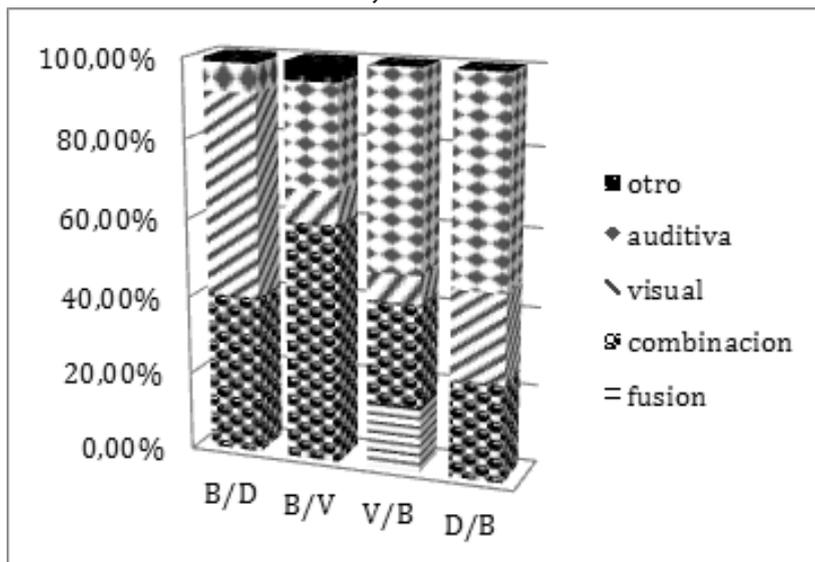


Gráfico 3 Porcentajes de tipo de respuesta según tipo de estímulo presentado en la serie de consonantes sonoras

El gráfico 3 muestra el porcentaje de cada tipo de respuesta seleccionado por los participantes en función de los 4 tipos de estímulos presentados. Para el estímulo B/D (consonante bilabial sonora en modalidad visual y consonante dental sonora en modalidad auditiva), el mayor número de respuestas corresponde a aquellas que consideran el estímulo visual sin confusión de sonoridad con un 51,19%. A continuación, se encuentran las respuestas de combinación con un 40,47%. En este estímulo las respuestas auditivas alcanzan un porcentaje muy bajo (7,14%).

En el caso del estímulo B/V (consonante bilabial como video, pero con una consonante velar sonora como audio), la mayor cantidad de respuestas se

concentra en las de combinación con un 60,71%. El número de respuestas auditivas sube en relación con el estímulo anterior y alcanza un 26,19%.

En el estímulo V/B (consonante velar sonora en la modalidad visual y una bilabial sonora en la modalidad de audio), el número de respuestas auditivas aumenta hasta el 50%. No obstante, aparecen algunas respuestas de fusión (16,66%) y, en mayor cantidad, las de combinación con un 26,19%.

El estímulo D/B (consonante dental como video y la consonante bilabial sonora en modalidad auditiva), es el que muestra menor uniformidad en las respuestas. El mayor porcentaje lo alcanzan las respuestas auditivas con un 52,38%.

3.2 Análisis de respuestas para consonantes sordas

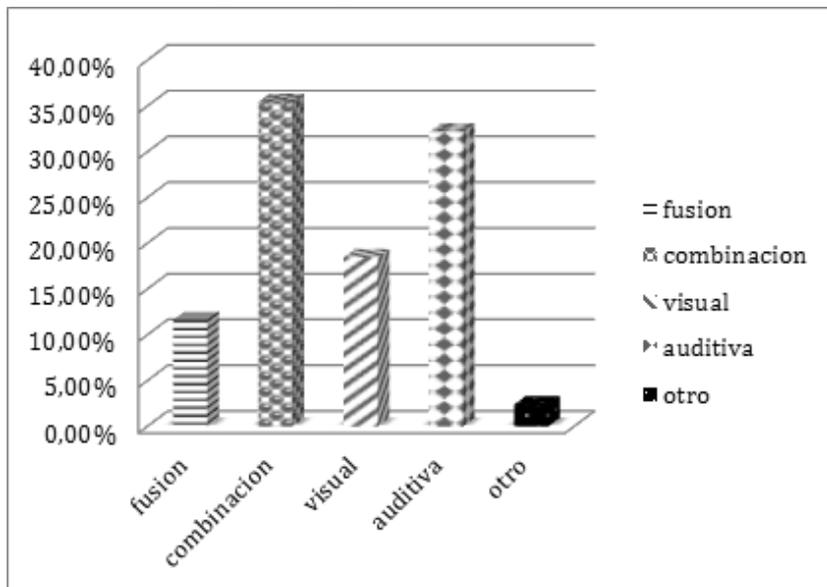


Gráfico 4 Porcentajes de los tipos de respuesta seleccionados en la serie de consonantes sordas

Según los datos obtenidos, representados en el gráfico 4, se puede apreciar que para las consonantes oclusivas sordas el patrón es muy similar al de las consonantes sonoras: las respuestas en que se produjo algún tipo de ilusión alcanzan el 65,47%; las respuestas en que no se produjo ningún tipo de ilusión obtuvieron un 32,14% y, por último, aquellas en que hubo confusión de sonoridad representan un 2,38% del total.

Si bien los porcentajes son bastante similares a los de la serie sonora, se

observa una diferencia importante: el porcentaje de respuestas de fusión aumenta a un 11,60%. Las respuestas de combinación representan un 35,41% del total y las visuales sin confusión de sonoridad un 18,45%.

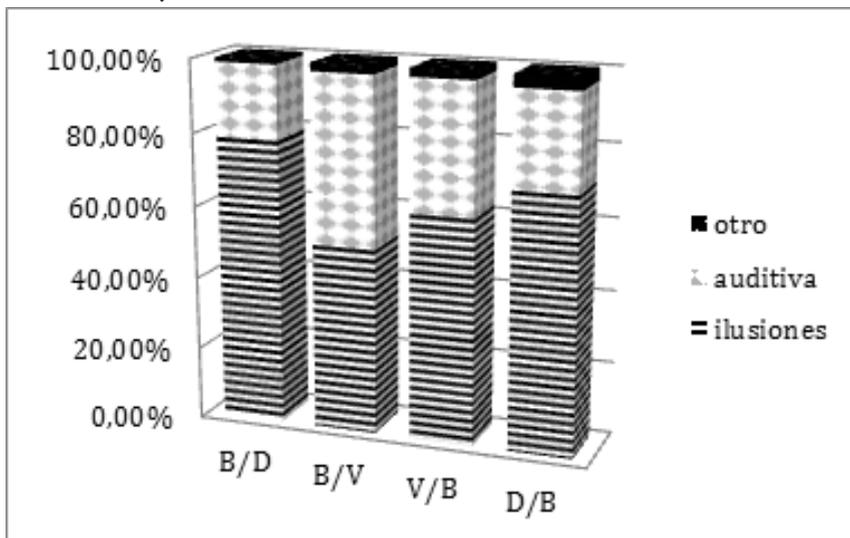


Gráfico 5 Porcentajes de ilusiones por tipo de estímulo presentado en la serie de consonantes sordas

El gráfico 5 presenta el porcentaje de ilusiones para los distintos estímulos presentados en la serie de consonantes sordas. En ella se puede observar que el mayor número de ilusiones también se concentra en el estímulo B/D (consonante oclusiva bilabial sorda en la modalidad visual y consonante dental sorda en la modalidad auditiva) con un 78,57%. El siguiente número elevado de ilusiones corresponde al estímulo D/B, en que la consonante dental es presentada como video y la consonante bilabial como audio, con un 70,23%. El estímulo V/B, de consonante velar como video y bilabial como audio, también presenta un número importante de ilusiones, alcanzando un 61,90%. El menor número de ilusiones se encuentra en el estímulo B/V (consonante bilabial presentada como video y velar en modalidad de audio). Es importante mencionar que, a diferencia de la serie sonora, en la serie de consonantes sordas el porcentaje de ilusiones en todos los casos supera el porcentaje de respuestas auditivas, es decir, en las que no hay ilusión.

Como se observa en el gráfico 6, para las consonantes sordas, en el estímulo B/D (consonante bilabial en modalidad visual y consonante dental en modalidad auditiva) la mayor cantidad de respuestas corresponde a combinaciones de ambas consonantes, como 'pta' o 'tpa' (40,47%). A continuación se encuentran las respuestas que consideran únicamente el estímulo visual, 'pa', con un 38,09%. Las respuestas auditivas alcanzan un

20,23%.

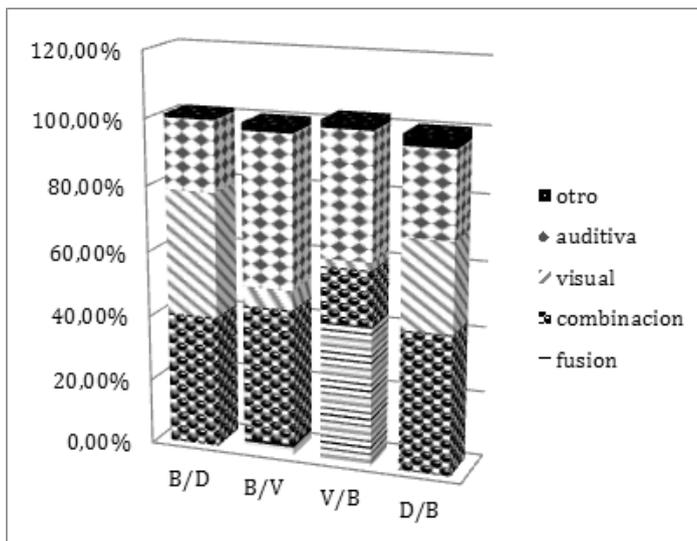


Gráfico 6 Porcentajes de tipo de respuesta por tipo de estímulo presentado en la serie de consonantes oclusivas sordas

En el estímulo B/V (consonante bilabial como estímulo visual con una velar como estímulo auditivo) la mayoría de las respuestas se concentra en las que consideran sólo el audio de 'ka' con un 46,42%. Las de combinación alcanzan un porcentaje levemente menor de 41,66%. Las respuestas visuales sólo alcanzan un 5,95%.

En el caso de V/B (consonante velar presentado en la modalidad visual junto con la bilabial en la modalidad auditiva) se puede apreciar que la mayoría de las respuestas corresponde al tipo fusión, alcanzando un 42,85%. Las respuestas de tipo auditivo presentan un porcentaje algo menor de 35,71%. Las de combinación alcanzan un 16,66%.

Finalmente, en la presentación del estímulo D/B (consonante dental sorda en modalidad visual y consonante bilabial sorda en modalidad auditiva), la mayor cantidad de respuestas corresponde a combinaciones con un 42,85%. Las respuestas visuales y auditivas alcanzan porcentajes similares de 27,38% y 26,19%, respectivamente.

4.3 Análisis de datos mediante pruebas de Chi-cuadrado

Luego de analizar los porcentajes obtenidos en el experimento, se presenta un análisis estadístico de las frecuencias de respuestas obtenidas mediante pruebas de Chi-cuadrado, las cuales permiten determinar si existe una

relación de independencia entre las variables propuestas: ‘tipo de respuesta’ y ‘tipo de estímulo presentado’.

Para trabajar los datos de la primera variable, ‘tipo de respuesta’, se reúnen los datos obtenidos (respuestas de los sujetos) en 5 categorías: auditiva, combinación, visual, fusión y otro. La primera (auditiva) corresponde a la percepción del estímulo auditivo. En la segunda (combinación) se agrupan las opciones que involucran ambos estímulos. La fusión, corresponde a la percepción de un estímulo que no es ni el visual ni el auditivo, sino uno intermedio. La respuesta visual es aquella que considera solo el estímulo visual sin confusión de sonoridad. Por último, la alternativa ‘otro’ corresponde a aquellas en que se considera sólo lo visual sin considerar la sonoridad del estímulo.

La segunda variable corresponde al tipo de estímulo creado; es decir, qué tipos de consonantes son utilizados en el estímulo presentado y en qué modalidades.

Las categorías de esta segunda variable son:

bA dV	(consonante bilabial Audio + consonante dental Video)
bV vV	(consonante bilabial Video + consonante velar Video)
dA bV	(consonante dental Audio + consonante bilabial Video)
vV bA	(consonante velar Video + consonante bilabial Audio)

En la prueba de Chi-cuadrado, se busca verificar la hipótesis nula entre ‘tipo de respuesta’ y ‘tipo de estímulo’. Además, para determinar si la sonoridad del estímulo presentado tiene algún efecto en la percepción, se añaden las capas ‘consonantes sonoras’ y ‘consonantes sordas’ en una segunda prueba de chi-cuadrado.

Como se explicó anteriormente, se somete a la prueba de Chi-cuadrado la hipótesis 0 (nula), según la cual la variable ‘tipo de estímulo presentado’ y la variable ‘tipo de respuesta’ no están relacionadas. La primera prueba considera el total de estímulos sonoros y sordos. Para esta prueba, tal como se aprecia en la tabla ii, se obtiene un valor observado del Chi-cuadrado igual a $\chi^2(1, N = 672) = 0.00$ $p > .05$. Es decir, la probabilidad de que las variables no estén asociadas es muy baja, por tanto se rechaza la hipótesis nula y se concluye que las variables están asociadas (H1).

Pruebas de Chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	272,204(a)	12	,000
Razón de verosimilitudes	266,623	12	,000
N de casos válidos	672		

4 casillas (20,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 3,25.

Tabla ii Prueba de Chi-cuadrado para tipo de estímulo y tipo de respuesta en consonantes sonoras y sordas

En otras palabras, estos resultados confirman que el tipo de estímulo presentado sí está relacionado con el tipo de respuesta que percibe el espectador. Al observar los residuos tipificados en la tabla de contingencia, se extraen conclusiones similares a las obtenidas en el análisis anterior. En la mayoría de casos, la frecuencia observada está por sobre lo esperado (supera el umbral de $\pm 1,96$). Cuando este número es negativo, como por ejemplo las respuestas auditivas en B/D, quiere decir que la interacción entre las variables genera un número menor a lo que se esperaría según la hipótesis nula. Al contrario, en el caso de las combinaciones y las respuestas visuales en B/D, este número supera significativamente lo que cabría esperar según la H_0 .

Para B/V la frecuencia observada supera la esperada en las combinaciones, respuestas auditivas y otro. En D/B esto sucede con los tipos de respuestas auditivo y otro. Finalmente en V/B, lo mismo sucede con las respuestas auditivas y con las de fusión.

La segunda prueba (tabla iii) se realiza añadiendo las capas 'consonantes sonoras' y 'consonantes sordas'. Para esta, el valor observado del Chi-cuadrado fue, nuevamente de $\chi^2(1, N = 336) = 0.00$ $p > .05$, tanto para las consonantes sonoras, como para las consonantes sordas. Esto quiere decir que se rechaza la hipótesis nula, según la cual las variables son independientes, tanto para la serie sonora, como para la serie sorda.

Pruebas de Chi-cuadrado

TipoCons	Valor	gl	Sig. asintótica (bilateral)	
Sonora	Chi-cuadrado de Pearson	150,383(a)	12	,000
	Razón de verosimilitudes	150,280	12	,000
	N de casos válidos	336		
Sorda	Chi-cuadrado de Pearson	157,392(b)	12	,000
	Razón de verosimilitudes	156,412	12	,000
	N de casos válidos	336		

8 casillas (40,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1,25.

4 casillas (20,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 2,00.

Tabla iii Prueba de Chi-cuadrado para tipo de estímulo y tipo de respuesta con capas ‘sonoras’ – ‘sordas’

Al observar los residuos tipificados en la tabla de contingencia, los valores que más se elevan sobre la frecuencia esperada para la serie sonora corresponden a la interacción entre el estímulo B/D y el tipo de respuesta ‘visual’; entre B/V y el tipo de respuesta ‘combinación’; entre D/B y el tipo de respuesta ‘auditiva’ y entre V/B y los tipos de respuesta ‘auditiva’ y ‘fusión’.

En la serie sorda las frecuencias observadas que más se elevan sobre las esperadas, corresponden a la interacción entre el estímulo B/D y el tipo de respuesta ‘visual’; entre el estímulo B/V y los tipos de respuesta ‘auditiva’ y ‘combinación’; entre el estímulo D/B y los tipos de respuesta ‘combinación’ y ‘visual’ y entre el estímulo V/B y el tipo de estímulo ‘fusión’.

En base a los datos obtenidos y analizados, tanto por medio de los gráficos como por las pruebas de Chi-cuadrado, se presentan a continuación algunas ideas que pueden ser un aporte a la discusión relacionada al fenómeno McGurk.

4. Discusión

En el detalle de los tipos de ilusiones generados por la serie sonora, presentados en el gráfico 3, se pudo observar que en el estímulo B/D el mayor número de respuestas corresponde a aquellas que consideran el estímulo visual (51,19%), seguidas por las respuestas de combinación con un 40,47%. En este estímulo las respuestas auditivas alcanzan un porcentaje de 7,14%. Es probable que este resultado se pueda explicar por la baja presencia acústica de la consonante dental y la fuerte presencia visual de la consonante visual. Es más, es el único estímulo en que el porcentaje de respuestas auditivas es tan bajo.

En el estímulo V/B, el porcentaje de respuestas auditivas es de 50%, mientras que el de las de fusión es 16,66% y de las de combinación 26,19%. En los estudios consultados el porcentaje de fusiones en las consonantes sonoras para este tipo de estímulo es de 52% en Behne *et al.* En Colin *et al.*, el número de fusiones en condiciones auditivas normales no supera el 17% en el caso de las consonantes sonoras para el mismo estímulo.

En dichos estudios la variable sonoridad lleva a distintas conclusiones según la lengua en cuestión. En el caso de Colin *et al.*, los hablantes del inglés perciben una mayor cantidad de combinaciones en los estímulos creados con consonantes sordas que con las sonoras. En el caso de las fusiones, si bien no alcanzan una significancia estadística, el porcentaje de fusiones es levemente mayor para las consonantes sonoras. McGurk y McDonald (1976) no encuentran diferencias importantes entre ambas series en el caso de la combinaciones, pero sí, al igual que en el caso del inglés, reportan mayor cantidad de fusiones para las consonantes sonoras. Sekiyama y Tohkura (1991), no obstante, reportan más fusiones para las consonantes sordas.

Esta diferencia entre ambas series de consonantes es explicada por Colin *et al.* Por una parte, la mayor cantidad de fusiones para las consonantes sonoras que reportan algunos autores, podría deberse a la mayor ‘confusabilidad en ruido’ que presentan estas. Lo percibido en una fusión se puede explicar como una consonante que sería fácilmente confundida, en presencia de ruido, con la consonante que efectivamente ha sido presentada como audio. Además, es compatible con la consonante presentada como video. De este modo, se podría pensar que la consonante bilabial sonora se confunde más fácilmente con la dental sonora, que en el caso de ‘p’ y ‘t’. El problema de esta explicación, señalan los autores, es que no resolvería el porqué las consonantes sordas generan más combinaciones.

Por otra parte, continúan los autores, se podría considerar que las fusiones

ocurren por una ambigüedad mayor tanto del estímulo auditivo como del visual. Por lo tanto, mientras más ambiguas las pistas presentadas, mayor sería la posibilidad de que se perciba un fonema fusionado. Según esto, las consonantes sonoras deberían producir más fusiones, por su menor saliencia acústica. En el caso de las combinaciones, en cambio, el mayor peso perceptual de las consonantes implicaría una percepción más nítida de estas. Por ello, la serie sorda debería generar más combinaciones que la serie sonora.

No obstante, y a pesar de que las conclusiones obtenidas en otros estudios parecen bastante razonables, los datos de esta investigación, basados en el español de Chile, arrojan resultados opuestos a los del inglés o el noruego. En la literatura del fenómeno McGurk, el par que generalmente se considera que conduce a las fusiones, es el de la bilabial sonora como componente auditivo y la consonante sonora velar como video (B/V), el cual lleva a la percepción de 'da'. En este estudio, dicho estímulo sí provocó algunas respuestas de fusión, pero un porcentaje bajo, como ya se ha mencionado anteriormente. Sin embargo, el mismo estímulo en la serie sorda, condujo a una mayor cantidad de respuestas de fusión, como se pudo observar en el gráfico 6.

En el detalle de las consonante sordas se puede apreciar que en el estímulo V/B, la mayoría de las respuestas corresponde a las fusiones con un 42,85%. Las respuestas de tipo auditivo representan un 35,71% las de combinación alcanzan un 16,66%. Esto confirma lo expuesto anteriormente: al contrario de los estudios que se han usado como referencia, las fusiones en el español de Chile se presentan en un mayor porcentaje en la serie sorda. Esto contradice la idea de que la ambigüedad acústica de las sonoras es la causa de que estas sean atraídas más fácilmente, generando así la percepción de una fusión. Si fuera así, además, las consonantes sonoras deberían conducir a menos combinaciones, pero como se puede observar en la comparación de los resultados de ambas series, son las consonantes sonoras las que arrojan un mayor porcentaje de combinaciones. Nuevamente, en el caso emblemático estudiado por otros autores (B/V), las combinaciones en la serie sonora superan en un 20% a la serie sorda.

5. Conclusiones

Al realizar el experimento dentro del paradigma McGurk, se puede comprobar que en el español de Chile también se presenta este fenómeno. Es más, el experimento realizado avala la teoría sobre este efecto, puesto que, en gran medida, los participantes responden ante los estímulos según lo estipulado por la teoría.

Luego de analizados los datos, tanto a través de los porcentajes obtenidos, como mediante el uso de pruebas estadísticas, es posible confirmar algunas ideas extraídas de otras investigaciones: existe efectivamente una relación entre cómo se presentan los estímulos auditivos y visuales, y lo que eso lleva a percibir. Además, se puede establecer que dicha relación es efectiva en ambas series de consonantes.

Siguiendo la literatura en torno al fenómeno McGurk, es posible afirmar también que se cumplen los tipos de ilusiones esperados según los tipos de estímulos presentados. Se puede determinar, además, que en el caso del par de la consonante bilabial y la consonante velar (el estímulo más utilizado en los estudios referidos), presentadas en los dos órdenes posibles, la frecuencia observada supera la esperada. Tanto para B/V, como para V/B, la interacción con los tipos de respuesta de combinación y fusión, respectivamente, muestra ser estadísticamente significativa. Más aún, la presente investigación permite corroborar que efectivamente ciertos sonidos consonánticos parecen tener más relevancia acústica que articulatoria, como en el caso de las velares, mientras que otros son percibidos mayormente por las pistas visuales que entrega el hablante al producirlos, como sucede con las bilabiales.

Un dato interesante extraído de esta investigación, es la baja confusión de sonoridad de los estímulos. Tanto para la serie sorda, como para la sonora, el porcentaje de respuestas con confusión de sonoridad, no supera el 5%. Esto, además de significar que los participantes en la mayoría de los casos hicieron uso tanto de lo visual como de lo auditivo, puede indicar que, en general, los hablantes del español de Chile no tienen problemas para reconocer un estímulo como perteneciente a la serie sonora o a la serie sorda. Tal vez, como señala Pérez (1998), esta diferenciación no tenga que ver con la ausencia o presencia de la barra de sonoridad, pero sí se puede afirmar que debe haber uno o varios rasgos que permiten a los sujetos identificar, casi en el total de los casos, una consonante sonora de una sorda. Al menos es así en el caso de las sílabas aisladas. Habría que explorar qué sucede en el caso de palabras, o incluso de frases. La dificultad sería generar estímulos en donde el componente semántico no influya el resultado. De este modo se podría indagar en el problema planteado por algunos autores sobre la sonorización de las consonantes en algunos dialectos del español.

Otra idea que es relevante destacar luego de estos resultados, es lo expuesto por Martínez y Fernández (2007) en cuanto a cómo la ausencia de un rasgo o lo no percibido puede tener un impacto importante en la percepción. Esto puede ser extrapolado a los resultados de esta investigación, en tanto que

la ausencia de una información visual puede conducir a una percepción distinta, como en el caso de la consonante velar como estímulo visual y la bilabial como estímulo auditivo.

Si bien con los resultados obtenidos no es posible resolver preguntas más fundamentales sobre cómo son procesados los estímulos auditivos y visuales, o sobre la especificidad del lenguaje, sí al menos se puede comprobar la integración audio-visual en estímulos incongruentes en el español de Chile. Esto indica que los usuarios de esta lengua, al menos en la mitad de los casos, hacen uso tanto de la información visual como de la auditiva. Es de suma importancia considerar que los porcentajes altos de respuestas auditivas en algunos de los estímulos creados, pueden deberse a factores propios de la realización de los estímulos y de la presentación de los mismos. Una edición profesional de los mimos podría tener como resultado el aumento en la cantidad de ilusiones. Asimismo, condiciones más apropiadas para llevar a cabo el experimento, como un estudio adaptado para dicha tarea, podrían tener un efecto en los resultados. Otro factor a considerar es el uso de audífonos de parte de los participantes. Tal vez en un futuro experimento sería más adecuado trabajar con parlantes o con el mismo audio del computador. Investigaciones anteriores han demostrado que mientras más degradado el sonido, mayor será el efecto.

A pesar de estas dificultades metodológicas, los resultados se encuentran dentro de lo esperado. Si bien los datos en relación al porcentaje de fusiones y combinaciones en consonantes sonoras y sordas, no coinciden con los obtenidos en investigaciones en otras lenguas, es interesante que estos reflejen un patrón opuesto al obtenido en ellas. Al contrario de lo que podría esperarse según algunos de los estudios antes nombrados, las consonantes sonoras presentaron un mayor número de combinaciones y un menor número de fusiones. Para la serie sorda, el patrón fue el inverso. No obstante, como se mencionó antes, existe al menos un estudio en donde se comprueba el mismo patrón que el obtenido en esta investigación. Esto pone en duda las afirmaciones hechas por otros autores y señaladas anteriormente según las cuales, la menor saliencia acústica de las consonantes sonoras debería generar más fusiones y menos combinaciones, al contrario de lo que sucedería con las sordas.

Sería pretencioso intentar esbozar una explicación absoluta para estos resultados. Por el momento se puede plantear la posibilidad de que esa mayor saliencia de las consonantes oclusivas sordas es lo que permite la percepción de un sonido fusionado más nítido. Por el contrario, la mayor ambigüedad que presentan las sonoras, hace que el resultado de la fusión de dos sílabas sea un sonido menos distinguible. En cuanto a las combinaciones, es posible

que la mayor saliencia acústica de las sordas inhiba el efecto que tiene la articulación de la bilabial sobre el sonido de la dental o de la velar; es decir, puede que la fuerza acústica de la consonante tenga un mayor peso perceptual en este caso. En las consonantes sonoras, en cambio, la menor saliencia del sonido de la oclusiva podría causar que lo visto y lo oído adquieran la misma fuerza en la percepción, lo que implicaría el uso de ambas pistas en una mayor frecuencia. No obstante, es preciso hacer notar que para el total de las combinaciones, la diferencia entre ambas series fue más estrecha que para las fusiones.

Con el objeto de intentar dar una solución más certera a los cuestionamientos ya planteados, sería preciso continuar con las investigaciones en esta área. Por ejemplo, parece importante indagar más en la diferencia de resultados entre diversas lenguas. Puede que las características de las consonantes de ambas series en distintos contextos lingüísticos, tenga una repercusión importante en la percepción del fenómeno. Sería interesante entonces hacer un estudio más acabado de las propiedades fonéticas de las consonantes empleadas en las otras investigaciones consultadas, para así poder realizar una comparación exhaustiva. Ciertamente quedan asuntos por explorar en cuanto al fenómeno mismo y a la percepción de los fonemas del español de Chile en general.

Bibliografía

- Baart, M. y Vroomen, J. (2010). Do you see what you are hearing? Cross-modal effects of speech sounds on lipreading. *Neuroscience Letters*, 471, 100–103.
- Behne, D., Wang, Y. Alm, M., Arntsen, I., Eg, R., & Valsø, A. (2007). *Changes in audio-visual speech perception during adulthood*. International Conference on Auditory-Visual Speech Processing
- Boersma, P. (2006). A constraint – based explanation of the McGurk effect. [en línea] <http://www.fon.hum.uva.nl/paul/papers/McGurk.pdf> (consultado en abril de 2011)
- Colin, C., Radeau, M., Deltenre, P., & Demolin, D. (2002). “The role of sound intensity and stop-consonant voicing on McGurk fusions and combinatios”. *European Journal of Cognitive Psychology* 475 – 491.
- Gentilucci, M. y Cattaneo, L. (2005). Automatic audiovisual integration in speech perception. *Brain Research* 167, 66–75.

- Gili y Gaya, S. (1961). *Elementos de fonética general*. Madrid: Gredos.
- Ladefoged, P. 1996. *Elements of acoustic phonetics*. Chicago: University of Chicago Press
- Martínez, E. y Fernández, A. (2007). *Manual de fonética española*. Barcelona: Ariel
- McGurk H. y MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–8
- Nicholls, M., Searle, D. y Bradshaw, L. (2003). Read My Lips: Asymmetries in the Visual Expression and Perception of Speech Revealed Through the McGurk Effect. *Psychological Science*, 15, 138-141.
- Omata, K. y Mogi, K. (2008). Fusion and combination in audio-visual integration. *Proceedings of the Royal Society*, 464, 319–340.
- Pérez, H. (1998). Incidencia de dos rasgos acústicos en la percepción de la correlación /p, t, k/ vs. /b,d,g/. *Revista de lingüística teórica y aplicada* 36, 113 – 126.
- Sekiyama, K. y Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America*, 90, 1797-1805.
- Skipper, J., van Wassenhove, V., Nusbaum, H. y Small, S. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17, 2387 – 2399.