

Pontificia Universidad Católica
Facultad de Letras
Departamento de Lingüística
Programa de Magíster en Letras mención Lingüística

**El efecto McGurk en consonantes oclusivas sonoras y sordas
del español de Chile:
una explicación fonético-acústico-perceptiva**

Tesista: Silvana Andrea D'Ottone Campana

Profesor guía: Domingo Román Montes de Oca

Índice

1. Introducción	4
2. Marco teórico	5
2.1 El efecto McGurk: historia y aplicaciones	5
2.2 Teorías de la percepción del habla	6
2.2.1 Teoría motora de la percepción del habla	7
2.2.2 Modelo lógico difuso de la percepción del habla	9
2.2.3 Teoría de la Optimalidad	10
2.3 La lectura de labios en la percepción audio-visual del habla	11
2.3.1 La lectura de labios: asociación y disociación de funciones	11
2.3.2 Integración audiovisual	13
2.4 Algunos estudios realizados en el paradigma McGurk	16
2.4.1 Rol de la sonoridad y de la intensidad en la percepción del efecto McGurk	16
2.4.2 Rol de la edad en la percepción del fenómeno	18
2.4.3 Escuchando labios y viendo voces: mediación de las áreas corticales en la producción del habla	20
2.5 Consonantes oclusivas	24
2.5.1 Las consonantes: descripción general	24
2.5.2 Las consonantes oclusivas sonoras y sordas	25
2.5.3 Las consonantes oclusivas del español	26
2.5.4 Las consonantes oclusivas del español de Chile	27
3. Objetivos	31
3.1 Objetivo general de la investigación	31
3.2 Objetivos específicos de la investigación	31
4. Metodología	32
4.1 Participantes	32
4.2 Procedimiento	32

5. Resultados	35
5.1 Análisis de respuestas en consonantes sonoras y sordas	37
5.2 Análisis de respuestas en consonantes sonoras	41
5.3 Análisis de respuestas en consonantes sordas	44
5.4 Análisis mediante pruebas de Chi-cuadrado	47
6. Discusión	51
7. Conclusiones	54
8. Bibliografía	58

9.

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 es presentar a un espectador el video 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.

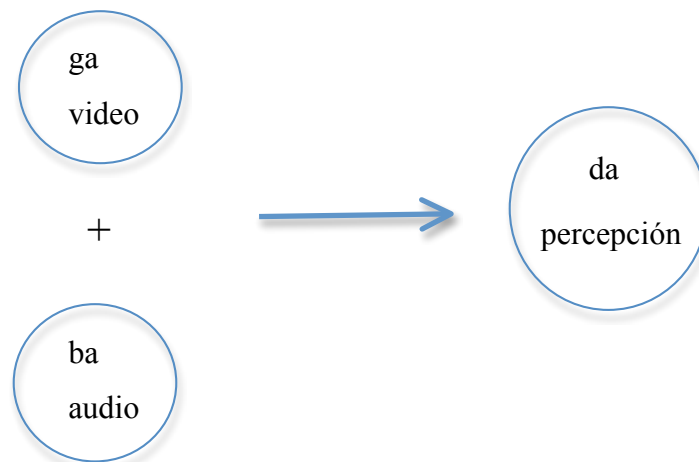


Figura 1 Esquema representativo del efecto McGurk

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 lo superficialmente perceptible) de las informaciones incongruentes y cómo estos son procesados por quien ve y escucha.

Dicho de otro modo, los rasgos acústicos de las consonantes, 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 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, 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 será principalmente fonética, pero se tomarán conocimientos provenientes de otras áreas.

Este estudio parte de una carencia en el ámbito de la fonética de nuestro país en el área de la percepción, a la vez que intenta entregar datos que puedan contribuir con la explicación del fenómeno mencionado. La presente investigación, entonces, podría contribuir en la descripción de los rasgos determinantes para la percepción de los fonemas oclusivos del español de Chile.

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 rasgos o marcas 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 hacer una revisión en torno a temas tales como percepción del habla, fonemas consonánticos oclusivos y, por supuesto, lo que se ha estudiado hasta el momento sobre el efecto McGurk.

2. Marco Teórico

2.1 El Efecto McGurk: historia y aplicaciones

Como se ha señalado antes, el efecto McGurk es un fenómeno de efecto fuerte bajo información

audio-visual 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.

Numerosos estudios han intentado, por un lado, dar a conocer las particularidades del fenómeno como, por ejemplo, el rol de la intensidad del sonido y de la sonoridad en la magnitud del efecto (Colin *et al.*, 2002); o bien, la importancia de la variable edad, o la variable cultural en la percepción de la ilusión (Behne *et al.*, 2007; Omata y Mogi, 2007). Por otro lado, el fenómeno se ha estudiado desde la psicología, en donde se han observado las respuestas conductuales de los sujetos a los estímulos incongruentes con el objetivo de obtener información sobre la percepción audio-visual. Algunos de estos estudios han intentado verificar la importancia de la significación fonológica para la percepción del fenómeno (Brancazio *et al.*, 2006) o la influencia del contexto semántico y la expectativa para rebatir la idea sobre lo “robusto” de la ilusión (Windmann, 2003).

Además, se han llevado a cabo diversas investigaciones en el área de la neurociencia y la neuropsicología, las cuales se basan en los efectos neurológicos del fenómeno para intentar dar una respuesta a la pregunta sobre cómo y dónde se produce la percepción de los fonemas (Gentilucci y Cattaneo, 2005; Skipper *et al.*, 2007; Baart & Vroomen, 2010). 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. Es por esto que antes de indagar más en estas investigaciones, es necesario entender las distintas posturas existentes en torno al tema de la percepción del habla, especialmente aquellas que parecen tener una mayor predominancia en la discusión sobre el efecto McGurk.

2.2 Teorías de la percepción del habla

La percepción del habla es un proceso complejo en torno al cual existen distintas posturas. Por

un lado, algunos autores, como Gòsy (1992) señalan que la percepción del habla es especialmente particular dentro de las percepciones humanas, porque es la única en la que el receptor puede producir el mismo tipo de señal que ha recibido. Por otro lado, quienes defienden una teoría de la percepción más amplia, afirman que no existe un módulo cerebral ni un proceso específico para el lenguaje.

Gòsy explica que todos los seres humanos poseerían una misma *base perceptual*, la cual define como “un set particular de operaciones mediante el cual el lenguaje hablado es procesado y en el cual las propiedades lingüísticas influyen de forma crucial la manera en que funciona el sistema” (11). Dicha base perceptual, además, sería independiente de la lengua materna; es decir, serviría a cualquier persona en cualquier contexto lingüístico. Sin embargo, que las personas posean estructuras anatómicas idénticas, señala la autora, no implica que estas tengan un mismo funcionamiento en los distintos sujetos. Cada lengua tiene su propia base perceptual que se acomoda en la adquisición de la misma y se relaciona con la base articulatoria.

También señala Gòsy que, dada la complejidad de estudiar cómo son procesados los sonidos del habla, han surgido una serie de *modelos de percepción del habla*, los cuales tienen como objetivo explicar y esquematizar los tipos de operaciones que se realizan en los distintos niveles del mecanismo de comprensión. Gòsy reconoce seis modelos, los cuales se distinguen por sus posturas teóricas, actitudes funcionales y sus objetivos finales; no obstante, todos ponen especial énfasis en la *especificidad del lenguaje*.

2.2.1 Teoría motora de la percepción del habla

Esta teoría, junto con la teoría motora revisada, de las cuales el mayor representante es Alvin Lieberman, han tenido gran repercusión fuera del área de la percepción del habla, como por ejemplo en la lingüística, al tiempo que ha sido duramente criticada entre los teóricos de la percepción. Básicamente, según esta teoría, existe un módulo especializado para los sonidos del lenguaje que convierte las señales acústicas en gestos articulatorios pretendidos por el hablante, lo cual conduce a las estructuras fonéticas percibidas. Es decir, el espectador no identifica los sonidos del habla, sino gestos abstractos del tracto vocal con los cuales se producen los sonidos: la señal de habla es analizada en términos de la propia producción de habla del oyente. Además, existiría un sistema motor específico e innato que se encarga tanto de la producción como del procesamiento del habla.

Como crítica ante este modelo y la especificidad lingüística que supone, Gòsy señala que la percepción de cualquier tipo de sonido (del habla o cualquier otro) implica los mismos mecanismos, independientemente de sus propiedades articulatorias; es decir, los distintos sonidos no son percibidos de forma distinta. Sin embargo, no niega la importancia de la articulación en la comprensión del habla.

Por otro lado, Tatham y Morton (2006) explican el modelo motor como un un proceso activo que involucra la cognición y la propia producción del oyente (los patrones articulatorios que pueden producir la señal). El conocimiento de las propiedades motoras y coarticulatorias del habla que posee el oyente, son parte de un proceso activo de decodificación de la onda sonora en etiquetas fonológicas apropiadas. El proceso coarticulatorio es evaluado por referencia a un contexto estático y lineal (no dinámico), en vez de a estructuras dinámicas jerárquicamente organizadas. El siguiente esquema intenta mostrar cómo se llevaría a cabo este proceso:

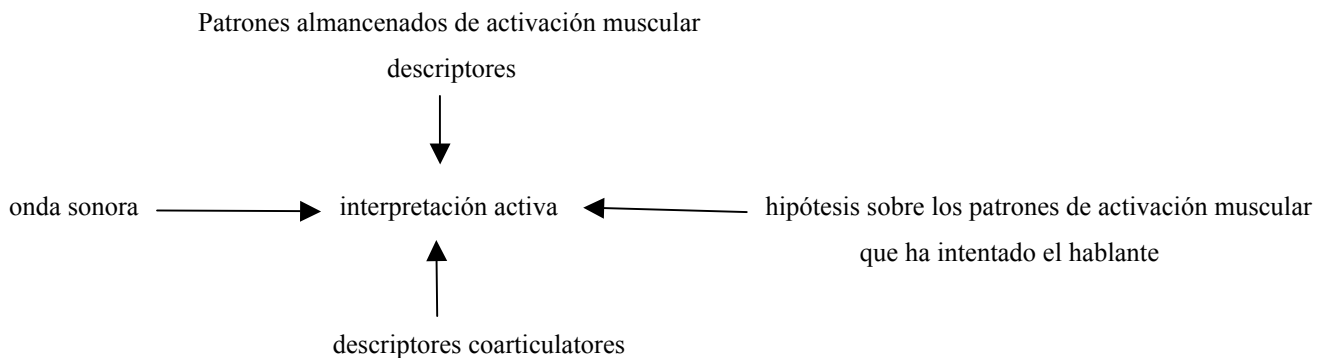


Figura 2 Proceso por el cual se logra la percepción según la teoría motora de la percepción del habla.

Extraído de Tatham y Morton (2006). Traducción propia

La *teoría motora revisada*, por su parte, plantea que en vez de patrones musculares, los hablantes de una lengua almacenan patrones gesturales, los cuales son sintetizados y reconstruidos por los receptores. Es decir, en esta versión del modelo la percepción no se basa en patrones de activación muscular invariantes, sino en gestos fonéticos. Estos gestos corresponderían a unidades abstractas, en base a las cuales los oyentes reconstruyen los planes de gestos articulatorios del hablante. Para esto, los receptores recurren a la información sobre sus propios procedimientos de producción.

Según esta teoría, la percepción no es visual ni auditiva; vale decir, no es bimodal, sino fonética

y es un resultado de la conversión temprana de los estímulos acústicos y visuales en los gestos articulatorios pretendidos por el hablante (Colin *et al*, 2002).

En contra de la teoría motora y de la teoría motora revisada, Tatham y Morton señalan, entre otras cosas, que ellas no pueden dar cuenta de aspectos dinámicos del habla como la prosodia o el contexto expresivo. Además, tampoco dan cuenta de la variabilidad.

A pesar de estas críticas, con el descubrimiento de las *neuronas espejo* la teoría ha ganado mayor popularidad entre los investigadores. Estas harían el vínculo entre la producción y la percepción de los movimientos motores, incluyendo los del tracto vocal. Las neuronas espejo constituyen un pequeño subconjunto de neuronas que originalmente se habrían encontrado en el área premotora izquierda F5 de los primates, la cual se activa durante la producción de acciones orientadas a una meta, pero también durante la observación de acciones similares. Un funcionamiento similar de este sistema espejo ha sido reportado en humanos.

Como explica Skipper (2007), existe evidencia conductual y neurofisiológica que defiende la idea de que el sistema espejo y, por ende, el sistema motor, jugarían un rol fundamental en la percepción del habla en un contexto de información audio-visual.

2.2.2 Modelo lógico-difuso de la percepción del habla (FLMP)

El modelo lógico-difuso o *fuzzy logical model of perception* (FLMP) nace de un sistema de lógica más general (lógica difusa), la cual reconoce que existen más valores que el de verdad y el de falsedad. Usando la conceptualización de la lógica, las proposiciones pueden ser representadas con grados de verdad o falsedad. La lógica difusa es útil para llegar a conclusiones definitivas a partir de sistemas complejos que generan información ambigua, vaga o imprecisa (Mathe, 2002).

El modelo lógico-difuso de percepción del habla (Massaro, 1987), supuso un nuevo enfoque en el área de la psicología. En este paradigma, se entiende que la información que se extrae de los eventos perceptuales es procesada de acuerdo con un algoritmo general y se realiza mediante tres momentos: la evaluación de rasgos, integración de rasgos y decisión. Como explica Mathe, según el FLMP, el sistema sensorial transduce (transforma) el evento físico y deja a disposición del oyente una serie de

fuentes de información llamadas *rasgos*, los cuales son evaluados en un *continuum*. Estos rasgos luego son evaluados y comparados con la descripción de los prototipos almacenados en la memoria del oyente. Luego, se lleva a cabo una decisión basada en lo correcto de la asociación de la información del estímulo con las descripciones relevantes del prototipo. El valor de “difusidad” asignado al rasgo predice la cantidad de veces que el estímulo es comparado con el prototipo. El impacto de una fuente de información aumenta en la medida en que aumenta la ambigüedad de las otras fuentes de información disponibles.

El FLMP es un modelo auditivo que da cuenta de la integración de fuentes de información que siguen una dirección ‘de arriba hacia abajo’ y ‘de abajo hacia arriba’ (*bottom-up/top-down*) en el procesamiento de los datos. Esto implica una integración tardía que se realiza después de la evaluación independiente y simultánea de las entradas acústica y visual. Esta idea contrasta con el modelo anterior, el cual apunta a una integración temprana. En este modelo, la integración se realiza de modo continuo y gradual. Además, el FLMP puede dar cuenta de la dependencia del contexto en la percepción, lo cual lo hace atractivo dentro de los modelos que explican la percepción del habla.

2.2.3 Teoría de la Optimalidad

Otro de los modelos que actualmente se están utilizando para explicar la percepción del habla es el de la Teoría de la Optimalidad. Este nace en la lingüística, especialmente en fonología, y básicamente indica que las decisiones de los hablantes para producir o percibir el habla se encuentran determinadas por una serie de restricciones jerarquizadas, las cuales evalúan las representaciones fonéticas y fonológicas y sus relaciones (Boersma, 2006). Es un modelo integrativo formal bidireccional de fonología y fonética

Las restricciones informan al hablante u oyente cuál es la salida o en la entrada más óptima entre una serie de posibilidades, en base a la jerarquía de dichas restricciones. Si una de las opciones viola una restricción de alto rango, esta será rechazada. Por lo tanto, de entre todas las posibilidades la opción elegida debe ser la más “óptima”.

2.3 La lectura de labios en la percepción audio-visual del habla

2.3.1 La lectura de labios: asociación y disociación de funciones

A partir de los estudios en torno a la percepción audio-visual, las investigaciones en torno a la *lectura de labios* han tomado mayor predominancia. Por un lado, debido a la importancia que supondría la incorporación del recurso visual para la percepción del habla en contextos de ambigüedad acústica, producto de situaciones de ruido en el ambiente; o bien, por déficits del oyente, por ejemplo, en casos de pérdida moderada de la audición. Por otro lado, el uso de las pistas visuales podría contribuir en el aprendizaje de segundas lenguas, especialmente de aquellas que poseen un sistema fonético que difiere en una medida importante al de la lengua materna.

Campbell (1992) explica que el uso del estímulo visual se ha reportado en infantes y sería, según la autora, un apoyo esencial en la sincronización del habla. Los individuos que se crían en un contexto de habla audio-visual podrían generar desde una edad temprana, la correspondencia entre el movimiento de los labios y el sonido vocálico, lo cual funciona como una fuente de información acerca de cómo se articulan los sonidos de su lengua. En este sentido, señala la autora, el fenómeno McGurk es una evidencia clara de que la percepción de la articulación puede modificar el sonido percibido. Además, ciertas aproximaciones teóricas apuntan a que las modalidades de entrada son críticas para la adquisición y el desarrollo de una habilidad particular.

Sin embargo, el tema parece tener gran complejidad para su estudio, especialmente desde la psicología o la neuropsicología, puesto que no se han reportado patrones claros de comportamiento respecto de la lectura de labios entre pacientes con diversos tipos de daño neurológico. Básicamente, la discusión se centra en qué áreas del cerebro son determinantes para una efectiva lectura de labios. Además, no se ha logrado dilucidar con claridad qué otras habilidades, como el reconocimiento de rostros o la lectura, son necesarias para desarrollar esta habilidad.

Campbell (1992), por su parte, da cuenta de una investigación en pacientes con distintos daños cerebrales. Los resultados de su investigación generan interesantes cuestionamientos. En esta, una de las participantes, la señora T., por ejemplo, presentaba un daño en el hemisferio izquierdo. Ella mostró habla y comprensión moderadas, pero serias dificultades para leer. Podía deletrear sin problemas, pero

no podía proyectar este conocimiento a las letras impresas (alexia). No parecía tener ningún problema visual, puesto que su habilidad para nombrar o reconocer objetos, caras, símbolos, era perfecta. La señora D. (lesión en el hemisferio derecho), en cambio, podía leer bien. Su habla y comprensión tampoco habían sido afectadas, pero no podía reconocer caras de personas que debían ser familiares para ella y no podía reconocer personas famosas. Era prosopagnósica. Además, tenía dificultades para reconocer las expresiones faciales. De las dos pacientes, la que presentó dificultades para leer labios fue la señora T., al contrario de lo que se habría pensado, puesto que no tenía otro problema para procesar rostros. La señora D., a pesar de que no podía reconocer rostros ni expresiones, podía combinar los movimientos de las caras con los sonidos correspondientes. También fue susceptible al efecto McGurk.

Estas pacientes, señala la autora, mostraron doble disociación de función. Estos resultados sugieren que el reconocimiento de identidades o expresiones, por un lado, y la lectura de labios, por otro lado, pueden usar módulos cognitivos distintos. Mientras que el hemisferio derecho puede ser crucial para los aspectos visuales, puede que no tenga un rol crítico en la lectura de labios. Tal parece que se necesitan algunas partes del hemisferio izquierdo para lograr una apropiada lectura de labios y para el procesamiento auditivo del habla. Estos resultados muestran que la lectura y la lectura de labios pueden o no coexistir, lo que sugiere separabilidad de procesos.

Otras pacientes estudiadas por Campbell, permiten suponer, además, que existe un nivel de análisis visual que sería fundamental para llevar a cabo la lectura de labios, el cual, sin embargo, puede ser independiente de otros procesos visuales como el reconocimiento de rostros o la lectura. Por otro lado, la habilidad de reconocer el movimiento de los labios parece también ser necesaria para desarrollar esta habilidad. En consecuencia, la lectura de labios podría ser un aspecto único del procesamiento facial: el movimiento de los labios no se caracteriza por un estado final o por una parte de la acción. Los patrones de cambio muscular constituyen el evento de la lectura de labios, y es el reconocimiento de estos patrones lo que permite una percepción efectiva de la articulación de los sonidos del habla.

En síntesis, la autora señala que a pesar de la disociación de funciones que presentan los pacientes, se pueden extraer algunos elementos que serían esenciales para el reconocimiento de los gestos articulatorios del habla: algún sitio de procesamiento del hemisferio izquierdo debe estar intacto para una lectura de labios efectiva; la percepción de eventos de movimiento también parece ser crucial;

finalmente, el análisis componencial sería determinante para la habilidad de lectura y de lectura de labios. No obstante, esto último no puede ser confirmado, lo que podría significar que ambas habilidades no funcionan mediante el mismo tipo de análisis.

2.3.2 Integración audio-visual

En el ámbito de la percepción audio-visual, Gentilucci y Cattaneo (2005) presentan un experimento usado para determinar si los rasgos tanto de las entradas acústicas como visuales, están siempre unidos en la representación del habla percibida y si esta integración audio-visual está basada en las funciones vinculantes a través de las modalidades (modelo lógico-difuso de la percepción), o en la imitación (teoría motora del habla). Para determinar esto, los autores trabajan dentro de un paradigma McGurk. En un primer momento, se solicita a los participantes repetir una cadena de fonemas, presentada en modalidad audio-visual. Los estímulos en este caso presentan congruencia de audio y video.

A continuación se pidió a los participantes que repitieran en voz alta otra cadena de fonemas emitida por un hablante en modalidad audio-visual. No obstante, en esta presentación, la cadena fonémica fue incongruente con la articulación del actor; es decir, se presentaron estímulos incongruentes. Los resultados fueron analizados en el nivel acústico y en cuanto a la cinemática de los labios. El primer análisis mostró tres patrones distintos: la fusión de los dos estímulos (efecto McGurk), repetición de la cadena acústica presentada y, menos frecuentemente, repetición de la cadena correspondiente a las modulaciones gestuales labiales. Sin embargo, el análisis de las respuestas en los dos últimos casos, mostró que el Formante 2 de los espectros de voz de los participantes, siempre fue distinto de los valores correspondientes a las repeticiones de las presentaciones audio-visuales congruentes (primera parte del experimento). En el análisis acústico de la repetición de los estímulos incongruentes, el formante 2 de lo emitido por los participantes (ya fuera la repetición de lo visto o lo oído, aparentemente), se aproximó a los valores del formante 2 de la cadena de fonemas presentada en la otra modalidad, la cual era aparentemente ignorada. Por otro lado, la cinemática de los labios de los participantes al repetir la cadena de fonemas presentada de forma acústica, fue influenciada por la observación de la articulación de los labios modulada por el actor, pero sólo cuando esta correspondió a una consonante labial.

De los datos extraídos en los experimentos, los autores señalan que el efecto McGurk no pudo ser comprobado sistemáticamente en hablantes de italiano, lo que podría explicarse por características

propias de esa lengua, la cual tendría un repertorio fonético y una realización fonética más simple que otras. En consecuencia, la identificación de fonemas no requeriría basarse tanto en las pistas visuales. Esto se puede relacionar con el modelo lógico-difuso de la percepción, según el cual la fuente menos compleja o menos ambigua de información, ya sea la visual o la acústica, es la más influyente para la percepción (Massaro, 1990)

Si bien no se reportó significativamente el fenómeno McGurk, es interesante que en el análisis espectral de voz realizado por los autores en la imitación de los participantes de los estímulos incongruentes, a pesar de que no existiera fusión, sí se pudo comprobar una influencia del estímulo presentado en la otra modalidad. El F2 (formante 2) de la consonante pronunciada por los espectadores se acercó al F2 de la consonante que supuestamente estaba siendo ignorada. Esto sugiere que sí hubo una percepción distinta de la cadena de fonemas, aunque esta no alcanzara a ser lo suficientemente fuerte como para percibir otro fonema. Mediante estos resultados y estas afirmaciones, Gentilucci y Cattaneo (2005) sugieren que el efecto McGurk se podría deber a la influencia de la modalidad visual, por ejemplo, en la percepción de la sílaba oída. Esta influencia cambiaría la percepción del F2, de manera que cambia el patrón de transición de formantes hasta ser percibido como otro sonido. Esto apoyaría la hipótesis de que la percepción audio-visual siempre contiene rasgos de las dos fuentes de información.

En cuanto al análisis cinemático, los autores pudieron comprobar que la percepción de la articulación de la consonante bilabial influyó la cinemática de los participantes al repetir la consonante velar; no obstante, esto no ocurrió en el caso contrario (consonante velar presentada como estímulo visual y repetición de la bilabial). Este resultado rechazaría la hipótesis sobre una superposición de programas motores de imitación automática de los patrones motores acústicamente presentados sobre los visualmente detectados. Por el contrario, señalan los autores, estos datos apoyan la hipótesis de una integración intermodal, la cual provee una información gradual y continua sobre las categorías del habla, en concordancia con la teoría de Massaro. No obstante, los autores no descartan que la imitación (teoría motora) sea utilizada en una etapa de filtro anterior a la integración intermodal.

Baart *et al* (2010) también investigan la integración audio-visual, pero de manera opuesta: intentan determinar si el habla oída afecta la lectura de labios, lo cual reflejaría una auténtica integración de las modalidades de las señales sensitivas. Además, intentan demostrar que la interacción entre lo oído y la lectura de labios es genuinamente bidireccional.

A partir de los resultados obtenidos, los autores proponen dos hipótesis de cuál podría ser la razón funcional para que haya una interacción entre el oír y el hablar. La primera es que es “ecológicamente” útil consultar más de una fuente, puesto que dos órganos sensitivos distintos proveen información complementaria sobre el mismo evento externo. Por esta razón, la lectura de labios es utilizada en la comprensión del habla puesto que puede compensar la interferencia producida por ruido externo y puede resolver ambigüedades internas de la señal auditiva. Una segunda razón es que hay “errores” internos entre los sentidos individuales que pueden ser ajustados por referencia cruzada a otras modalidades. En el dominio del habla, este tipo de referencia cruzada a otras referencias ha sido reportada solo en el último tiempo y consistiría en un ajuste perceptual inducido por una desviación entre dos fuentes de información que el cerebro intenta reducir. El estudio extiende estos descubrimientos mostrando que este tipo de ajuste no sólo ocurre para el habla oída, sino también para el habla “vista”.

Los descubrimientos de la investigación, señalan Baart *et al*, también son relevantes en el ámbito de los mecanismos neuronales involucrados en el procesamiento multisensorial del habla audio-visual. Imágenes neuronales y estudios electrofisiológicos han descubierto interacciones audio-visuales en áreas multimodales como el Surco Temporal Superior (STS) y áreas sensoriales específicas incluyendo las cortezas visuales y auditivas. Se ha propuesto que las entradas unimodales (auditiva y visual) se integran inicialmente en el STS y que las interacciones en las cortezas visuales y auditivas primarias reflejan una retroalimentación entre estas y el STS. Según esta lógica, las interacciones en la corteza primaria probablemente estén mediadas por el STS, a través de “proyecciones hacia atrás” o “retroproyecciones” (*backward projections*).

Además del STS, otras regiones motoras de planificación y ejecución (área de Broca, corteza premotora, e insula anterior), podrían estar involucradas a través de las llamadas neuronas espejo. Se ha propuesto que el área de Broca sería homóloga a la corteza premotora inferior de algunos primates, en la cual se ha descubierto este subconjunto de neuronas que se activan ante la acción y/o percepción de movimientos de manos o de labios orientados hacia una meta. Presumiblemente, la acción de estas neuronas es mediar la imitación y la comprensión. Los autores, señalan que el área de Broca no sólo estaría involucrada en la producción del habla, sino que también se activaría durante la lectura de labios silenciosa y la atención pasiva del habla. De ser así, la activación de las neuronas espejo en el área de Broca podría facilitar el vínculo entre el habla oída y vista, junto con sus correspondientes

representaciones motoras.

De esto se concluye, entonces, que la visión puede afectar el procesamiento auditivo a través de los programas motores articulatorios de los actos de habla observados, y como dicen haber demostrado los autores, es probable que la naturaleza de este efecto sea bidireccional.

2.4 Algunos estudios realizados en el paradigma McGurk

2.4.1 Rol de la sonoridad y de la intensidad en la percepción del efecto McGurk

Ahora, ya más específicamente 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. Basándose en investigaciones anteriores, los autores intentan determinar si existen diferencias en la magnitud del fenómeno cuando se presentan consonantes sonoras y sordas, ya sea en la modalidad visual o auditiva. Además, en relación con estudios anteriores que investigan el fenómeno bajo condiciones auditivas degradadas, como sucede en la adición de ruido, la investigación tiene como objetivo determinar si existe una relación entre el porcentaje de ilusiones reportadas y la intensidad de la señal acústica. Para comprobar esto presentan estímulos de 70 dB y 40 dB.

Los resultados tanto de las pruebas de control y las pruebas experimentales, muestran que entre los participantes (hablantes de inglés), el número de errores al reconocer consonantes fue similar en las condiciones de sólo audio e incongruente. Este resultado ser de un 1% para las consonantes bilabiales y de un 20% para la consonantes velares. Los autores descartan que la pista visual de la boca cerrada tuviera influencia en los resultados en el caso de las bilabiales, puesto que en la prueba control, en la cual sólo se presentaron estímulos acústicos, el porcentaje de error fue muy similar: 0% para las bilabiales y 16% para las velares.

En concordancia con otras investigaciones en torno al efecto McGurk, los autores pudieron determinar 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. Es importante señalar que para rechazar la hipótesis de que los participantes se basaron sólo en los estímulos visuales en la condición de audio

degradado, los autores eliminaron aquellas respuestas que presentaron confusión de sonoridad; es decir, en el caso de que ante la presentación de /b/ como estímulo visual y /g/ como estímulo auditivo, el sujeto perciba /p/, se considera que no existe ilusión, ya que la persona sólo está considerando la articulación vista y no el componente auditivo. En el caso de las ilusiones, el estímulo auditivo otorga a la persona la información de la sonoridad.

Otro dato interesante arrojado por esta investigación es que el número de fusiones fue 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 presentaron un mayor número de combinaciones que las sonoras y, si bien no fue estadísticamente significativo, el número de fusiones fue un tanto mayor para las consonantes sonoras. Este mismo patrón fue reportado por McDonald & McGurk (1978) y Sukiyaama & Tohkura (1991).

Una posible explicación que plantean los autores se relaciona con la “confusabilidad” acústica en presencia de ruido, en el caso de las fusiones. Según Summerfield (1987), la percepción audio-visual producto del efecto McGurk, es una consonante que sería fácilmente confundida, en presencia de ruido, con la consonante presentada en la modalidad auditiva en el estímulo creado y que, además, es compatible con la consonante presentada en la modalidad visual (/da/ en el caso del estímulo incongruente A/b/ V/g/). El hecho de que este efecto se reporte más comúnmente con consonantes sonoras podría deberse a que /b/ es más ambigua que /p/ y tiende a confundirse con /d/, mientras que /p/ no es confundida con /t/, en el caso del inglés. Esto explicaría por qué la bilabial sonora da pie a más fusiones que la sorda.

Otra posible explicación tiene que ver con la “saliencia” perceptual de las consonantes, en términos acústicos y visuales. Bajo esta lógica, la obtención de una ilusión u otra dependería del mayor o menor peso perceptual de la información acústica o visual. En términos visuales, la saliencia es mayor para las bilabiales que para las velares o las dentales. En términos acústicos, la energía de la explosión es mayor para las consonantes velares, debido a que en ellas el punto de oclusión es más posterior en la cavidad bucal.

De este modo, las combinaciones se explicarían por una fuente de información visual muy fuerte de las consonantes bilabiales, la cual no puede ser ignorada, junto con la información acústica de las

consonantes velares, la cual también tendría un importante peso perceptual. Las fusiones, por su parte, serían explicadas por la información visual ambigua de las consonantes velares y la información fonética de las bilabiales, la cual también tendría poco peso perceptual. Esto llevaría a una percepción intermedia entre lo visto y lo oído.

Ahora bien, el hecho de que las consonantes sordas den lugar a más combinaciones, sería explicado, señalan los autores, porque la explosión de la consonante velar sorda /k/ posee un mayor peso perceptual que la de la sonora /g/. Por otro lado, la explosión es menos intensa en las consonantes sonoras lo que permitiría que estas fueran “atraídas” por la consonante en la modalidad visual, produciéndose así las fusiones.

Si bien estos resultados parecen concluyentes, hay que considerar que las características de los fonemas del inglés no son equivalentes a las de los fonemas del español, por lo que se pueden usar estos datos como referencia, pero es probable que los resultados de esta investigación lleven a otro tipo de conclusiones. En un estudio preliminar hecho en el español de Chile, se ha observado que el patrón de fusiones es inverso a lo propuesto por Colin *et al.* Esto deberá ser explicado en su momento en base a las características acústicas de los fonemas del inglés en contraste con los rasgos acústicos del español de Chile.

2.4.2 Rol de la edad en la percepción del fenómeno

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 audiovisuales en la identificación del lugar de articulación del estímulo percibido.

Para neutralizar las diferencias en la habilidad auditiva entre ambos grupos, los autores utilizan dos fondos de ruido: ruido de café (ruido de fondo) y sin ruido. Esto les permitió rechazar la hipótesis de que las diferencias en la percepción se debieran a una compensación al cambio natural en la percepción por la edad. Los resultados obtenidos permitieron determinar que las respuestas de los participantes no se vieron afectadas por el tipo de fondo de ruido utilizado. Esto confirmaría que las diferencias entre los grupos etarios de 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 pudo observar también en este estudio, que las consonantes sonoras condujeron más a las fusiones que las consonantes sordas. Para la identificación del lugar de articulación de las sordas, los participantes habrían usado las pistas visuales y auditivas de manera independiente.

Los autores pudieron concluir, además, que las pistas visuales se utilizan más en la condición de mayor ruido, pero, siempre y cuando, estas ya estaban en uso en la otra modalidad (sin ruido). Es así como los adultos de mediana edad presentaron mayor tendencia a usar las pistas visuales en ambas condiciones, en contraste con los adultos jóvenes, quienes mostraron un uso menor de ellas en los dos ambientes de ruido.

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, indican 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 indicaban los resultados de Behne *et al.* (2007), lo cual además se relaciona con la experiencia articulatoria. Bajo esta lógica, serían apropiados los planteamientos de la Teoría Motora, según la cual existiría una relación cercana entre percepción y producción.

Los resultados obtenidos en este estudio, son consistentes con la idea de que el aprendizaje de la percepción AV juega un rol importante en el proceso que lleva al efecto McGurk. Por otro lado, hay evidencia de que los niños recién nacidos muestran actitudes consistentes con la integración audio-visual. Por ende, es probable que el aprendizaje de la percepción AV no genere el efecto McGurk, sino que lo cultive.

Es interesante destacar que los datos de dicha investigación coinciden más con un modelo de integración temprana que con un modelo de integración tardía (FLMP), lo cual lo situaría en una posición más cercana a las teorías motoras de la percepción audio-visual.

Ahora, si bien existen estudios que apoyarían los postulados de la teoría motora, estos se encuentran lejos de tener la última palabra en el tema, en especial porque también existe una serie de otros estudios cuyos resultados abogarían más bien por una postura cercana al modelo lógico difuso. Otros estudios en el área de la neurociencia y la neuropsicología, postulan una hipótesis más bien intermedia, según la cual tanto los programas motores como la integración tardía de las modalidades visual y auditiva estarían involucrados en la percepción del habla, solo que en distintos momentos de este proceso. Esto, unido al problema de la plasticidad del cerebro para cumplir las funciones de procesamiento sensorial, lleva a deducir que probablemente no exista una única respuesta frente al problema, sino ciertas tendencias o patrones que pueden indicar, en mayor o menor medida, cuáles son los pasos o momentos que deberían conducir a la percepción audio-visual.

2.4.3 Escuchando labios y viendo voces: mediación de las áreas corticales en la producción del habla

El estudio de las neuronas espejo, ha llevado a los investigadores a proponer que el hecho de observar el movimiento de los labios de una persona al hablar, da pie a un plan motor en el oyente, el cual sería usado para poder producir el movimiento observado. Como se explicó anteriormente, las neuronas espejo son un pequeño subconjunto de neuronas, encontradas en el área premotora, que se activa ya sea durante la producción de una acción dirigida a un objetivo, o bien, durante la observación de acciones similares. Evidencia neurofísica y de comportamiento apoyan la noción de que el sistema espejo humano y por tanto el sistema motor, juegan un rol crítico en la percepción del habla cuando se

observan los movimientos de la boca.

Con el uso de resonancia magnética funcional, se ha demostrado previamente que la percepción audio-visual del habla activa una red de áreas motoras, incluyendo el cerebelo y las áreas motoras corticales involucradas en la planificación y ejecución de la producción del habla. También se ha mostrado que son los aspectos visuales, y no los auditivos, los que activan este sistema motor.

En esta misma línea de pensamiento, el estudio de Skipper *et al* (2007) confirma la noción de que la percepción audio-visual del habla está mediada, en parte, por la actividad cortical, la cual es producto de experiencia previa del sujeto en la producción del habla. En otras palabras, el reconocimiento del habla audio-visual involucra comandos motores que podrían haber producido los movimientos de boca observados.

Sin embargo, lo anterior no es suficiente para explicar cómo el cerebro “entiende” que la actividad en el sistema motor es relevante para la percepción del habla. Para ilustrarlo, los autores recurren al modelo de percepción de habla basado en “hipotetizar-y-probar” o “análisis-por-síntesis”. Según este modelo, representaciones multisensoriales tempranas del habla, que derivan de patrones de sonido y movimientos faciales observados, pueden ser pensados como hipótesis multisensoriales, mas no interpretaciones finales, sobre los fonemas producidos por el hablante. Estas hipótesis, extraídas de la información sensorial, son mapeadas en comandos motores usados en la producción del habla. Estos comandos motores son asociados con hipótesis basadas en experiencia pasada sobre producción de habla. Luego, estos comandos motores activados predicen las consecuencias acústicas y somatosensoriales de ejecutar un movimiento de habla, mediante una copia eferente. Estas consecuencias generadas internamente pueden restringir la interpretación fonética de la información sensorial de entrada mediante, por ejemplo, dar peso a una interpretación particular de un fragmento de un enunciado que corresponde a la representación sensorial del fonema estadísticamente asociado con la producción del fonema hipotetizado.

Los resultados del estudio muestran, en primer lugar, que algunas áreas corticales activadas durante la producción del habla, se activan tanto en la percepción de habla audio-visual congruente como incongruente, y que esta actividad ocurre principalmente cuando los movimientos de la boca están siendo observados por el oyente. En segundo lugar, se mostró que la clasificación activa de los estímulos McGurk y McDonald en dos sílabas diferentes por parte de los oyentes (algunos perciben

/ka/ y otros /ta/ ante la presentación del audio de /p/ y la articulación de /k/) genera dos patrones distintos de actividad en las áreas corticales frontales involucradas en la producción del habla. Del mismo modo, las distintas sílabas congruentes, al ser vistas y oídas, producen diferentes patrones de actividad en dichas áreas. En tercer lugar, se pudo determinar que el patrón de actividad generado por la sílaba incongruente en las áreas corticales frontales involucradas en la producción del habla, se asemeja mucho al que produciría la sílaba audio-visual congruente correspondiente a la percepción de los participantes (/ka/ o /ta/). Y por último, se pudo observar que en un primer momento de la respuesta hemodinámica, el patrón de actividad generado por la sílaba McGurk en las áreas sensoriales, se asemeja al patrón de actividad producido por la sílaba audio-visual congruente correspondiente a la entrada sensorial (el estímulo auditivo usado para crear la sílaba incongruente). No obstante, en un segundo momento se asemeja al patrón de actividad producido por la sílaba audio-visual que corresponde a la última percepción del participante.

Los autores interpretan el primer resultado como una prueba de que existe un mecanismo compartido de producción y percepción, en el cual la producción de sonidos distintos requiere la coordinación de diferentes grupos de músculos, o de los mismos grupos pero en proporciones distintas. Este sistema motor estaría involucrado en la percepción fonética final del estímulo audio-visual, puesto que el mismo estímulo percibido de distintas maneras es asociado a dos patrones distintos de actividad.

También plantean que estos resultados demostrarían que la actividad en las áreas del sistema motor asociado con la producción del habla durante la observación de habla vista y oída, representa una hipótesis acerca de una interpretación fonética particular de las propiedades del estímulo, más que una representación verídica o apropiada de las propiedades de ese estímulo. La respuesta en el sistema motor, asociado a la producción del habla, corresponde a la última percepción de los participantes.

Por lo anterior, los autores concluyen que la activación del sistema motor no es la percepción en sí. El sistema motor refleja la actividad que corresponde a la última percepción de la sílaba, mientras que los sistemas auditivo y somatosensorial cambian de un patrón de actividad basado en lo sensorial, a uno que es más consistente con la actividad en el sistema motor frontal. Por lo tanto, en vez de interpretar la actividad del sistema motor como una imitación o ensayo, o como lo que constituye la percepción, los autores interpretan los resultados según el modelo antes descrito. Este modelo indica que la actividad en el sistema motor asociado a la producción del habla, constituye una hipótesis, la cual es usada para predecir las consecuencias acústicas y somatosensoriales de ejecutar esa hipótesis a

través de una copia eferente.

Otra conclusión fundamental del estudio, se relaciona con la teoría motora de la percepción y la teoría de análisis-por-síntesis. Los autores sugieren que el proceso de percepción del habla involucra interacciones entre un conjunto distribuido de córtices sensoromotoras, mediante mecanismos de procesamiento de proyecciones hacia delante y proyecciones hacia atrás o retroproyecciones (*feedforward* y *feedback*). La teoría motora, explican los autores, sugiere que la percepción del habla ocurre directamente por referencia a programas motores invariantes de producción del habla. De este modo, todo habla es conducido directamente a un código gestual. Dichos autores también plantean que no hay un procesamiento auditivo del habla al servicio de la percepción de fonemas. Skipper *et al.* plantean, no obstante, que a pesar de que los movimientos faciales asociados al habla generan una actividad motora cortical en regiones del cerebro involucradas en la producción del habla, la percepción del habla no sólo está determinada por un mecanismo directo de proyección hacia delante (*feedforward*). Por el contrario, los resultados del estudio sugieren que las entradas sensoriales interactúan con copias eferentes (retroalimentación) del sistema motor involucrado en la producción del habla. Estos resultados ponen en cuestionamiento la naturaleza de la interacción entre las cortezas motoras y sensoriales en la percepción del habla, proponiendo que el flujo de la información no es unidireccional, sino que se transmite de atrás para delante de manera constante.

Como se puede extraer de esta revisión, el tema de la percepción del habla es un asunto que aún no encuentra una solución definitiva y que conduce a constantes discusiones entre investigadores que adhieren a diversas posturas. El efecto McGurk ha resultado ser un fenómeno interesante en esta materia, puesto que entrega nuevas pistas sobre el funcionamiento del proceso de percepción del habla, tanto oída como vista.

Ahora, del mismo modo como se pueden encontrar diferencias entre las formas de entender el procesamiento del habla, también en el ámbito de fonética y fonología existen distintas posturas, entre otras cosas, sobre la naturaleza de las consonantes y, sobre todo, sobre cuáles son aquellos rasgos que permiten su identificación o percepción. A continuación se presentarán algunas propuestas en relación a las consonantes oclusivas en general, y más específicamente, aquellas que constituyen el repertorio de consonantes oclusivas del español de Chile.

2.5 Consonantes oclusivas

2.5.1 Las consonantes: descripción general

Las consonantes oclusivas pueden definirse como sonidos consonánticos en donde “los órganos activos de la articulación establecen un contacto completo que obstruye momentáneamente la salida del aire por la boca” (Gili Gaya, 1961). Este proceso articulatorio consta de tres fases: la fase implósiva, en que los órganos activos se ponen en movimiento para adoptar el gesto articulatorio de contacto correspondiente; la segunda fase oclusiva que corresponde al momento en que 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).

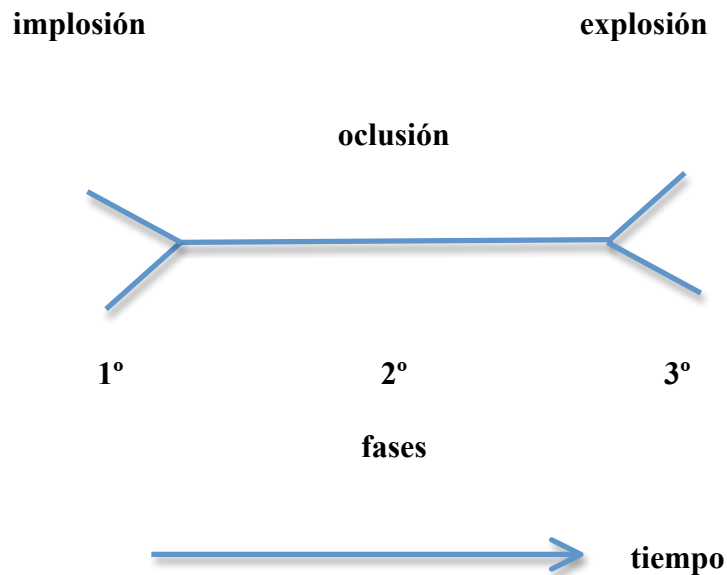


Figura 3 Fases de una consonante oclusiva. Extraído de Martínez y Fernández (2007)

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).

2.5.2 Las consonantes oclusivas sonoras y sordas

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 implosión 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. El movimiento del primer formante sólo indica que existe un cierre (en el paso del aire). Para las tres consonantes oclusivas sonoras, la frecuencia del primer formante aumenta al comienzo de la sílaba y declina hacia el final. Por lo tanto, el F1 no diferencia a estas consonantes. Sin embargo los formantes 2 y 3 diferencian a una de otra. Si el segundo y tercer formante incrementan rápidamente de frecuencia, entonces se trata de una /b/. Si el tercer formante cae y el segundo tiene poco movimiento, se trata de una /d/. Si los formantes 2 y 3 están juntos, la parte trasera de la lengua está en contacto con el paladar, como en /g/.

Las consonantes oclusivas sordas son hechas con los mismos movimientos de labios y de lengua que las tres anteriores. Por lo tanto, el movimiento de los formantes es similar para ambos conjuntos de sonidos. La diferencia entre estos sonidos es la vibración de las cuerdas vocales: mientras que en las oclusivas sonoras existe vibración de la cuerdas vocales, en las sordas no. En consecuencia, al inicio de una sílaba con /p/, /t/ o /k/, existe una explosión de aire que produce un tipo distinto de sonido. En vez del sonido producido por la acción de las cuerdas vocales, el cual tiene un *pitch* o tono definido, en las oclusivas sordas se produce un sonido más ruidoso y sin un pitch o tono, dado que no hay periodicidad.

2.5.3 Las consonantes oclusivas del español

Para Martínez y Fernández (2007), la perceptibilidad de los sonidos del habla se da gracias a una relación entre elementos que, en el caso de las consonantes oclusivas u obstruyentes, involucra tanto lo audible de la fase explosiva del sonido, como lo que no es percibido (ausencia de sonido), pero sí marca un contraste en el contexto auditivo. 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. El oído percibiría la interrupción de la emisión fónica, lo cual tendría un gran valor informativo. Por el contrario, la primera fase implosiva no tendría un correlato para la percepción. Finalmente, la tercera sería el momento entre la explosión y la vibración de las cuerdas vocales; sin embargo, los estudios realizados por el autor muestran que la explosión no es necesaria para la percepción de una consonante oclusiva, aunque sí tal vez para distinguir el punto de articulación.

Desde esta perspectiva, en términos articulatorios son esenciales la primera y la segunda fase; es decir, aquellas en que los órganos articulatorios se unen y permanecen en contacto. Acústicamente, resultan relevantes las tres fases, ya que todas ellas tienen algún tipo de correlato en el espectrograma. Perceptivamente, la segunda y la tercera serían las fases esenciales puesto que permiten el contraste entre la ausencia de sonido y la presencia del mismo.

Uno de los rasgos que se ha utilizado frecuentemente para distinguir a las consonantes es el de sonoridad. De manera casi unánime en la literatura sobre consonantes, se pueden diferenciar dos series de oclusivas: las consonantes oclusivas sordas /p t k/ y las consonantes oclusivas sonoras /b d g/. Acústicamente, señalan los autores, la sonoridad se presente en el espectrograma como un formante que corresponde a los pulsos glotales.

Otro elemento vinculado con la sonoridad es el V.O.T. (voice on set time) el cual corresponde a la medición del tiempo que transcurre entre la barra de explosión del espectrograma y el comienzo de los pulsos glotales (barra de sonoridad). Si la barra de sonoridad comienza antes que la explosión, se trataría de una consonante sonora cuyo V.O.T. será negativo; mientras que si la barra de sonoridad es posterior a la explosión, la realización corresponderá a una consonante sorda con V.O.T. positivo.

Finalmente, otro elemento que puede resultar interesante para esta investigación es la distinción entre tenso y laxo, a pesar de que no ha sido bien definido. Según el autor, el rasgo tenso se define fisiológicamente por “su mayor fuerza articulatoria y espiratoria, lo cual tiene como consecuencia su mayor duración relativa” (69).

De lo anterior se desprende un debate entre los investigadores sobre la pertinencia del rasgo de sonoridad. Especialmente en el español, explica el autor, tal parece que el rasgo 'sordo' no se presenta en el habla cotidiana. Por este motivo, Martínez y Fernández, junto con otros investigadores del área, están de acuerdo en considerar las oclusivas sordas como tensas y las sonoras como laxas. Esta distinción se da principalmente por la duración, índice principal de la tensión.

Las investigaciones en torno a la sonoridad y la tensión, han llevado al autor a concluir que sonoridad no es fundamental para distinguir entre las dos series de oclusivas. Por el contrario, la modificación en la duración de las consonantes al parecer sí podría cambiar la percepción de un sonido. Además, la sonorización de las consonantes en varios dialectos del español, y que también parece estar muy presente en el español de Chile, apoya los argumentos del autor.

2.5.4 Las consonantes oclusivas del español de Chile

En cuanto a las características de las consonantes oclusivas del español de Chile, no existe una cantidad importante de investigaciones. Uno de los estudios reconocidos en torno a este tema es el de Soto-Barba & Valdivieso (1999). En él se intenta determinar cuáles son los rasgos que diferencian los grupos consonánticos, sonoras y sordas, desde una perspectiva fonético-acústica.

Los rasgos acústicos observados fueron: el V.O.T, la duración absoluta (desde el inicio hasta el término de la consonante), el IREDUS (índice relativo a la duración de la sílaba), la intensidad de la onda periódica y la velocidad de las transiciones vocálicas.

Los resultados obtenidos indican que, según los parámetros observados, el V.O.T., es decir, el tiempo de inicio de la sonoridad, diferencia notablemente las dos series de consonantes: en las sordas, los valores se alejan de cero (valores negativos) a medida que la consonante se articula hacia la

zona posterior de la cavidad bucal; en la serie sorda, los valores se acercan a cero a medida que la consonante se articula hacia la zona posterior de la cavidad bucal.

También la duración absoluta de las consonantes diferenciaría ambas series consonánticas: La serie sorda muestra valores casi tres veces más altos que el grupo sonoro. Al parecer, esto también estaría relacionado con la zona de articulación, puesto que en la serie sorda la duración aumenta a medida que la articulación es más posterior, mientras que en las sonoras ocurre lo contrario.

Si bien el IREDUS también diferencia ambas series, no presenta exactamente el mismo patrón que los rasgos acústicos anteriores. En este caso, la serie sorda presentó valores que doblan el de las sonoras. Este valor, además, aumentó en la medida en que la articulación de la consonante se realizó en un punto más posterior de la cavidad bucal. Sin embargo, el patrón inverso no fue demostrado.

Finalmente, la intensidad de la onda periódica permitió determinar que la ausencia total de ella en las consonantes sordas, sería otro rasgo diferenciador de ambas series.

En base a estos datos, los autores concluyen que la diferencia entre ambos grupos se debe a una combinación de rasgos y no a alguno o algunos en particular, por lo que ante la ausencia de uno de estos rasgos, la distinción entre sonido sonoro y sordo puede ser mantenida.

A raíz de esta investigación se puede observar que existen discrepancias en cuanto a la caracterización de las consonantes oclusivas. Una de ellas guarda relación con el concepto de sonoridad, puesto que, como se señaló anteriormente, la distinción sonoro/sordo presenta una serie de dificultades teóricas y es parte de una larga discusión entre los investigadores. Una de las principales razones es que se cuestiona que en algunos dialectos del español exista esta distinción en el habla cotidiana. Algunos estudios del español y del español de Chile han demostrado una acrecentada tendencia a la sonorización de los fonemas tradicionalmente conocidos como sordos. Esta tendencia parece no estar ausente en el habla de nuestro país, y si bien sólo es una hipótesis preliminar, tal parece que esta distinción se ha ido perdiendo cada vez más en el habla espontánea.

Ahora bien, como se mencionó antes, el estudio descrito fue realizado desde una perspectiva acústica, mientras que la presente investigación tiene una orientación perceptiva. Por supuesto, no son miradas excluyentes y de alguna manera lo articulatorio, lo acústico y lo perceptivo tienen que estar

presentes en este estudio. No obstante, es muy probable que algunos de los rasgos que acústicamente diferencian a las consonantes no sean fundamentales para la discriminación que hacen los oyentes. Esto podría explicar, entre otras cosas, por qué los espectadores del efecto McGurk pueden percibir un sonido que acústicamente ‘no existe’.

Ahora, en el área de la percepción propiamente tal, Pérez (1998) es uno de los pocos estudios en el español de Chile. Según este, los estudios fonéticos-descriptivos que se han llevado a cabo en Chile no responden la pregunta sobre la incidencia de los rasgos observados en la discriminación de sonidos oclusivos sonoros o sordos.

El autor parte de la base de que algunos rasgos, que acústicamente pueden tener valores y características definidas, sencillamente no tienen un correlato en la percepción de los mismos. La tensión, por ejemplo, es un rasgo problemático porque tiene múltiples manifestaciones acústicas y no sería susceptible de ser estudiada. Las transiciones, por su parte, son un rasgo especialmente errático en sus manifestaciones acústicas, por lo que su estudio tampoco es efectivo. El V.O.T. y el I.R.E.D.U.S. finalmente, son sólo unidades de medidas y no constituyen rasgos acústicos.

En vista de lo anterior, la duración es uno de los rasgos netamente acústico y se puede medir en milisegundos. La sonoridad, a su vez, es susceptible de ser modificada mediante síntesis, a través de la adición o supresión de su manifestación acústica en el espectrograma. En consecuencia, el estudio se centra en estos dos rasgos. Básicamente, lo que se intenta es manipular estos rasgos que sí son claros y que sí se pueden alterar, con el objetivo de comprobar si, al hacer esto, los participantes experimentan o no un cambio en la percepción de los sonidos sonoros y sordos.

Según los datos obtenidos, se pudo concluir que la sonoridad, dada por la ausencia o presencia de barras de baja frecuencia, no incidiría en la discriminación de las consonantes sonoras y sordas. Por el contrario, la duración absoluta de la consonante sí tendría influencia para distinguir las consonantes oclusivas bilabiales y las consonantes oclusivas dentales de ambas series. Al igual que Soto-Barba & Valdivieso, el autor sugiere que la discriminación de consonantes sonoras o sordas, probablemente responde a una interacción de rasgos

Luego de esta revisión se puede constatar que los estudios del efecto McGurk han tenido una mayor predominancia en el área de la percepción y, por tanto, en disciplinas como la psicología o la neuropsicología. Como ya se ha explicado, esto se debe al gran potencial del fenómeno para estudiar aspectos de la percepción audio-visual que no habían podido ser explorados con experimentos de estímulos congruentes. Si bien la lingüística también ha jugado un rol importante en estas investigaciones, no se ha reportado una cantidad considerable de estudios propiamente lingüísticos, ni tampoco alguno que entregue explicaciones en base a los rasgos fonéticos y fonológicos de los estímulos empleados. Sin ir más lejos, no se han encontrado estudios en español del efecto McGurk, fuera de algunas referencias menores en trabajos con otra finalidad.

En Chile actualmente existe escasa investigación en el área de la percepción del habla, al menos desde la lingüística. Tal como se explicó antes, 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. El efecto McGurk, no obstante, es una herramienta probada que, si bien aún presenta muchas incógnitas, parece ser una buena posibilidad para contribuir a la descripción perceptiva de los fonemas oclusivos del español de Chile.

A pesar de los límites que pueda tener esta investigación, se espera que los datos obtenidos en ella sean un aporte al estudio de la percepción fonética del español de Chile. Además, este estudio dará a conocer un fenómeno muy interesante, el efecto McGurk, el cual no sólo es relevante para el estudio de la percepción, sino que también puede ser utilizado para la investigación en distintas áreas como, por ejemplo, en trastornos del lenguaje o en aprendizaje de segundas lenguas.

3. Objetivos

3.1 Objetivo general de la investigación:

Realizar un aporte al estudio de la percepción de los fonemas oclusivos sonoros y sordos del español de Chile (/b/, /d/, /g/ - /p/, /t/, /k/)

3.2 Objetivos específicos de la investigación:

Comprobar el fenómeno McGurk en el español de Chile.

En el caso de comprobar este efecto en el español de Chile, describir las respuestas de los hablantes ante los estímulos presentados.

Esbozar una explicación de dichas respuestas, considerando las características acústicas y articulatorias de los estímulos empleados en las dos modalidades.

Comparar los resultados con los obtenidos en otras investigaciones del efecto McGurk en distintas lenguas.

4. Metodología:

A continuación se darán a conocer los materiales y procedimientos empleados para llevar a cabo la investigación y, con ello, intentar resolver los objetivos planteados.

4.1 Participantes:

El experimento se llevó a cabo con 84 sujetos (18-25 años) hablantes nativos de español de Chile que no presentaran algún tipo de impedimento auditivo o visual (en este último caso, si se tratara de un problema corregido, el sujeto será igualmente válido para el procedimiento). Para verificar que se cumplan estas condiciones de este tipo, los participantes fueron consultados antes de participar en el experimento.

4.2 Procedimiento:

El experimento consiste en la presentación de estímulos audio-visuales, incongruentes, los cuales muestran un hablante de género femenino (sólo 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 hizo en el programa Imovie del sistema operativo Macintosh. El experimento constó de 8 estímulos:

Serie sonora:

baV daA

baV gaA

daV baA

gaV baA

Serie sorda:

paV taA

paV kaA

taV paA

kaV paA

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 Pyscope del sistema operativo Mac OS X. En este se presentaron las instrucciones para realizar la tarea. Luego, los estímulos se presentaban en modalidad de pantalla completa. A continuación de cada estímulo, se desplegaban las 7 opciones. El participante debía pulsar en el teclado la letra correspondiente a la alternativa que creyera haber percibido. Una vez que se realizara una elección, el programa arrojaba el siguiente estímulo. Las respuestas de los sujetos eran grabadas por el programa.

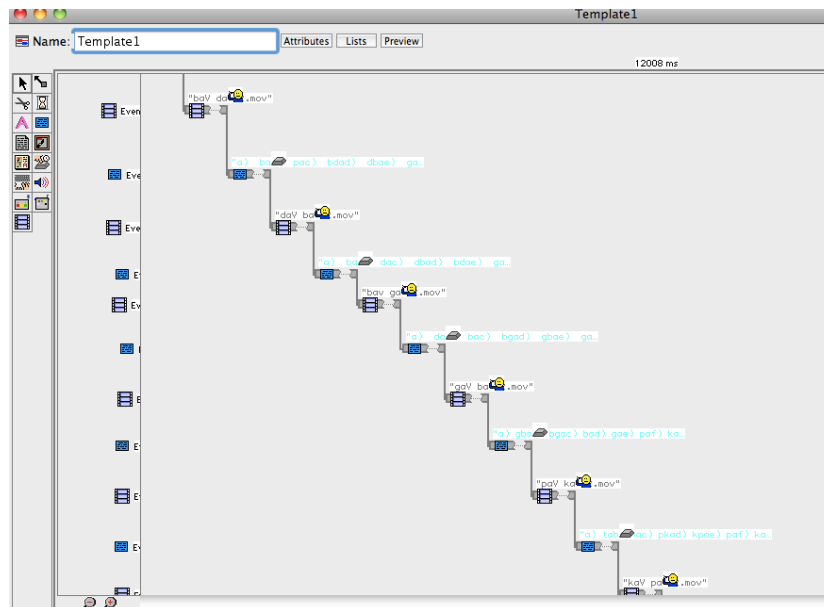


Figura 4 Diseño del experimento creado en el programa Pyscope de Mac OS X



Figura 5 Imagen representativa del tipo de video utilizado durante el experimento

Para la realización del experimento, cada sujeto fue situado frente a la pantalla de un computador Mac de 13 pulgadas. Además, se le entregaron audífonos. De este modo, los participantes podían ver y oír el estímulo simultáneamente. Luego de la aparición de cada estímulo, el sujeto debía elegir lo que creía 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 se 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

La primera opción corresponde al estímulo visual. En este caso, se considera que se trata de una ilusión, ya que la persona incorpora el movimiento de los labios como información articuladora y la

información acústica, la cual le informa de la sonoridad.

La segunda opción corresponde sólo a la información acústica. En este caso, se estima que la persona no incorpora, o bien, ignora la información articuladora, por lo cual no se puede considerar como un proceso audio-visual.

Las opciones tres y cuatro, 'bda' y 'dba', corresponden a combinaciones entre las informaciones acústica y articuladora. Es decir, el sujeto incorpora ambos estímulos en la percepción.

La opción 'ga' es lo que los estudios del efecto McGurk han denominado 'fusión'. En este caso, los individuos procesan ambas modalidades, acústica y visual, y perciben un fonema que no corresponde a ninguno de los estímulos presentados.

Las dos últimas opciones contienen las consonantes correlativas de la serie sorda. Si el sujeto percibe una de estas sílabas, se considera que no ha habido una integración audio-visual, puesto que no se está incorporando la información auditiva, la cual debería indicar que el estímulo es sonoro y no sordo. Colin *et al.* denominan a esto *confusión de sonoridad*.

Las opciones que aparecían luego de cada estímulo fueron ordenadas al azar.

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.

5. Resultados

Los datos obtenidos fueron analizados, en un primer momento, de forma descriptiva mediante la lectura de los gráficos elaborados a partir de los mismos.

En un segundo momento se llevó a cabo la prueba estadística de Chi-cuadrado, la cual pone a prueba la hipótesis de nulidad (H0) según la cual dos variables no están relacionadas; es decir, las variables observadas son independientes. En el caso de esta investigación, se intentó determinar si

existe una relación de independencia entre las variables 'tipo de respuesta' y 'tipo de estímulo presentado'. Según esta prueba, si la significación estadística de la prueba no es suficiente como para asumir una relación de dependencia de las variables, se considera que estas son independientes (las variables no están relacionadas). Por el contrario, si el valor observado del Chi cuadrado excede el umbral determinado por los grados de libertad, se tiene que aceptar que las diferencias entre las frecuencias esperadas y las frecuencias observadas no pueden ser atribuidas al azar, lo que obliga a concluir que las variables están asociadas de manera estadísticamente significativa.

Para trabajar los datos como variables, se reunieron las respuestas en 5 grandes tipos: auditiva, combinación, visual, fusión y otro. La primera corresponde a la percepción del estímulo auditivo. En la segunda 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. Esta primera variable se denominará 'tipo de respuesta'.

La segunda variable corresponde al tipo de estímulo creado; es decir, qué tipos de consonantes han sido utilizados, y en qué modalidades, en el estímulo presentado. En este caso, las categorías 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 buscó verificar la hipótesis nula (las variables no están relacionadas) entre 'tipo de respuesta' y 'tipo de estímulo'. En el caso de que esta no se comprobara, se asumiría que existe algún tipo de relación entre la forma en que se combinan las consonantes en el estímulo creado, y el efecto que esto tiene en la percepción del mismo por parte del espectador.

Además, para determinar si la sonoridad del estímulo presentado tiene algún efecto en la percepción, se añadieron las capas 'consonantes sonoras' y 'consonantes sordas' una segunda prueba de chi-cuadrado.

Es importante señalar que en las investigaciones consultadas, los pares de consonantes utilizados son diversos y han sido determinados en base a las características de las distintas lenguas en estudio. De todos modos, el caso más emblemático del fenómeno McGurk se da con el par velar-bilabial. En este estudio se decidió trabajar, además, con el par dental-bilabial. El par dental-velar fue eliminado de la investigación debido a la similitud auditiva y visual de ambos fonemas.

Para el análisis descriptivo, los resultados se han agrupado de tres formas: en un primer momento, 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 para las consonantes sonoras y sordas (total de respuestas de los participantes); luego, sólo con la serie sonora y, finalmente, con la serie sorda.

Como se explicó anteriormente, para efectos de este análisis, se considerarán 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 sólo ha considerado la información del movimiento de los labios, pero no la información auditiva que debería informarle acerca de la sonoridad del estímulo.

5.1 Análisis de respuestas en consonantes sonoras y sordas

El análisis de las respuestas, tal como se explicó antes, se centra en primer lugar en la totalidad de respuestas de los participantes; es decir, lo que se observa en la figura 6, corresponde a los porcentajes de respuestas seleccionadas por los participantes tanto en el caso de las consonantes oclusivas sonoras, como de las consonantes oclusivas sordas.

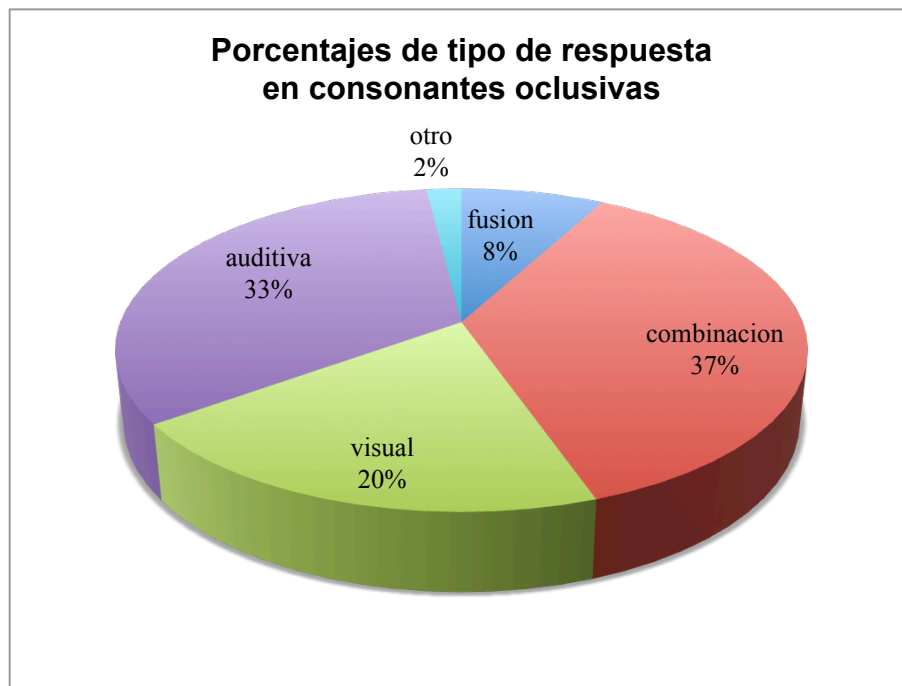


Figura 6 Gráfico de porcentajes de los tipos de respuesta seleccionados para ambas series de consonantes oclusivas

Para el total de estímulos creados, tanto con consonantes oclusivas sonoras, como con consonantes oclusivas sordas, el tipo de respuestas que obtuvo un mayor porcentaje fue aquél que implica la combinación de las dos consonantes presentadas en la modalidad visual y en la auditiva, con un 36,75%. Las respuestas en que los participantes consideraron sólo el estímulo auditivo, alcanzaron un porcentaje levemente menor de 33,03%. Las de tipo visual sin confusión de sonoridad alcanzaron un 20,38%. Las de fusión representan un 7,88% del total. Las respuestas visuales en donde sí se considera que hubo confusión de sonoridad ('otro'), es decir, que no se tomó en cuenta el estímulo auditivo, sino sólo el movimiento de los labios, obtuvieron un 1,93%.

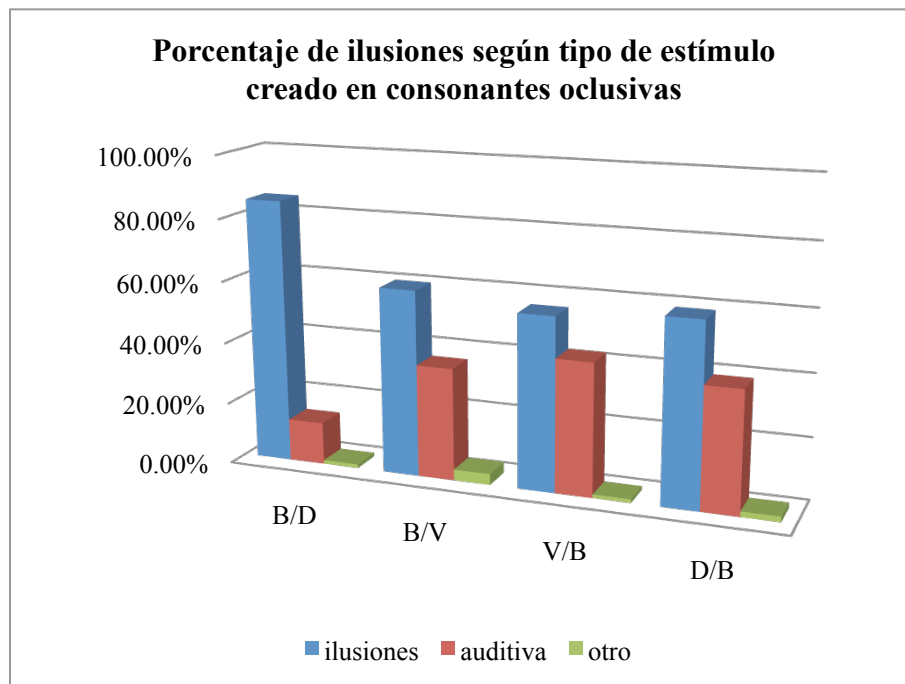


Figura 7 Gráfico de porcentaje de ilusiones según tipo de estímulo creado en ambas series de consonantes

La figura 7 muestra las respuestas de combinación, fusión y visuales, agrupadas bajo el rótulo “ilusiones”. Con ello, es posible determinar el porcentaje de respuestas que corresponden a ilusiones. Para el total de estímulos, tanto de consonantes oclusivas sonoras como de consonantes oclusivas sordas, la mayor cantidad de respuestas fue de aquellas que corresponden a algún tipo de ilusión, ya sean de combinación, fusión o visuales. El número de ilusiones obtenido con el experimento alcanza un 65,09%. Las respuestas en que los participantes consideraron sólo el estímulo auditivo, es decir, en las que no hubo ilusión, obtuvieron un 33,03% del total. En último lugar se encuentran aquellas en donde existe confusión de sonoridad y que han sido catalogadas bajo el rótulo “otro”. estas sólo alcanzaron un 1,93%.

Para el total de respuestas de los participantes en las ocho pruebas, es decir, considerando las consonantes oclusivas sonoras y sordas, se observa que el tipo de estímulo que generó más ilusiones es el compuesto por una consonante bilabial en modalidad visual, y una consonante dental en modalidad auditiva (B/D). Para el resto de los estímulos, el porcentaje de ilusiones es similar entre ellos y se eleva en los 4 casos por sobre el 50% del total de respuestas de los participantes. En estos tres estímulos, las respuestas auditivas fluctúan entre el 30 y 40% del total de respuestas.

Al observar las respuestas de los participantes en el par bilabial-velar se puede determinar que, si bien el porcentaje de ilusiones es más alto que el de respuestas auditivas, es menor que en el par bilabial-dental y las respuestas auditivas también aumentan.

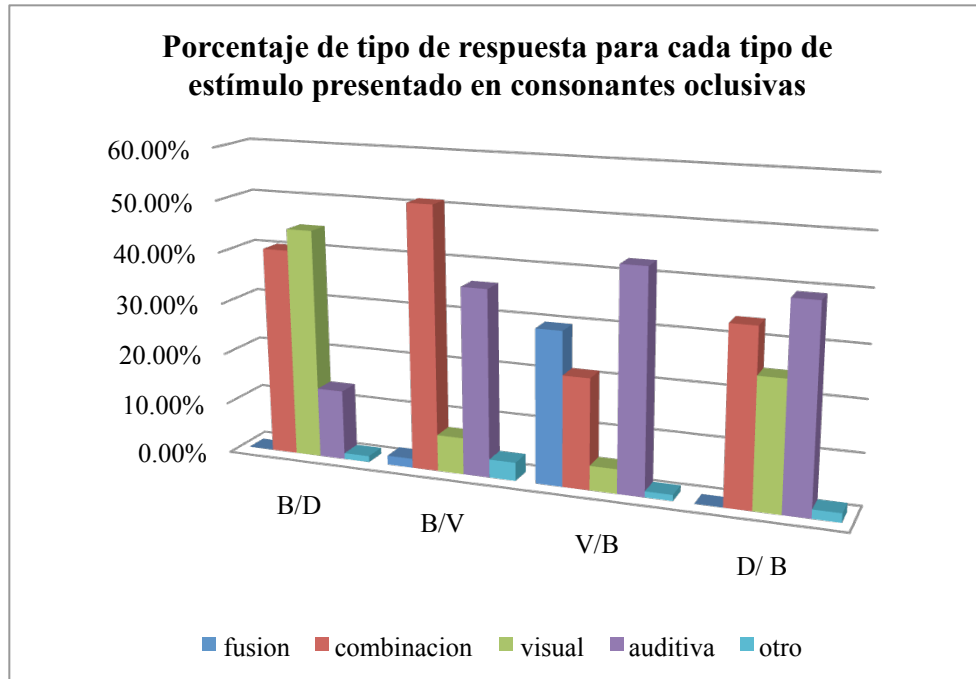


Figura 8 Gráfico de porcentajes de tipo de respuesta según tipo de estímulo presentado en ambas series de consonantes

Si se analizan los datos de la figura 8, en la cual se distinguen los diferentes tipos de ilusiones percibidas por los participantes en función de los estímulos creados, se puede extraer que para el estímulo de consonante bilabial en modalidad visual junto con la consonante dental en modalidad auditiva (B/D), la mayor cantidad de respuestas fue de tipo visual (44,64%), y en un porcentaje levemente menor, se encuentran las de combinación (40,47%).

Para el estímulo de consonante bilabial en modalidad visual y consonante velar en modalidad auditiva (B/V), la mayor cantidad de respuestas corresponde a combinaciones de ambos estímulos, alcanzando un 51,19%. Las respuestas auditivas se presentan en un porcentaje menor de 36,30%

En el estímulo constituido por la articulación de una consonante velar y el audio de una consonante bilabial (V/B), la mayoría de las respuestas se concentran en las que consideran sólo el

aspecto auditivo (42,85%). Sin embargo, las de fusión alcanzan casi un 30%, seguidas por las de combinación con un 21,42%.

En el caso del estímulo constituido por una consonante dental en el video y una consonante bilabial en el audio (D/B), parece haber menores diferencias perceptuales entre tres tipos de respuesta: las visuales, las de combinación y las auditivas. Si bien estas alcanzan el mayor porcentaje, 39,28%, las siguen las de combinación con un 33,92% y las visuales con un 25%. Esto podría significar que la menor uniformidad en las respuestas de los participantes puede deberse a que la consonante dental, tanto auditiva como visualmente no tienen una presencia importante en relación a los otros dos tipos de consonantes.

5.2 Análisis de respuestas consonantes sonoras

A continuación, se presentan los datos obtenidos para la serie de consonantes sonoras.

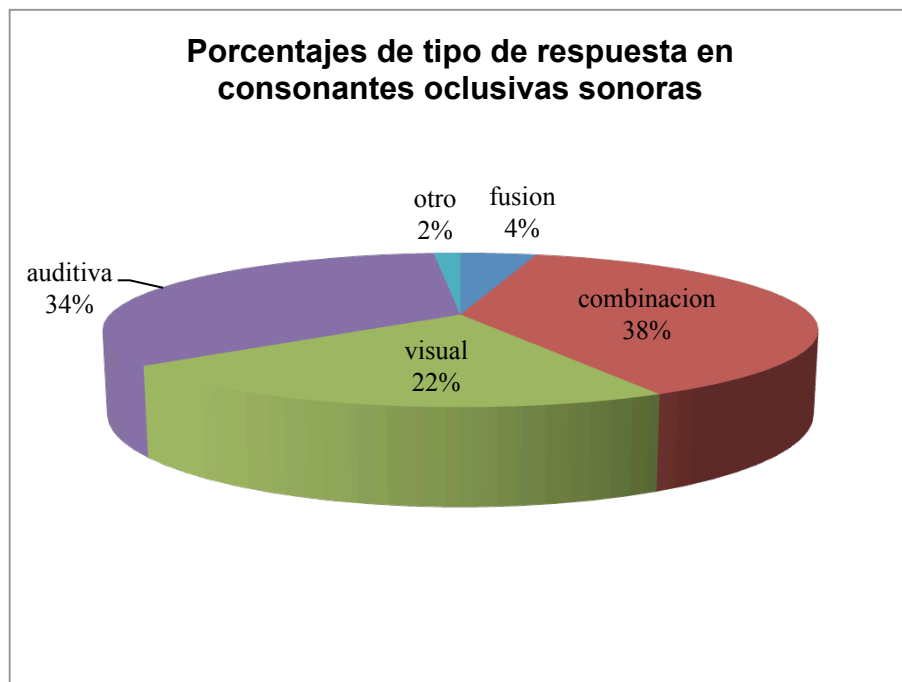


Figura 9 Gráfico de porcentajes de los tipos de respuesta seleccionados en la serie de consonantes sonoras

Los resultados obtenidos para las consonantes oclusivas sonoras, expresados en la figura 9, muestran que la cantidad de respuestas que implican algún tipo de ilusión, corresponde al 64,58% del

total. Las respuestas que consideran sólo el estímulo auditivo obtuvieron un porcentaje de 33,92%. Aquellas en que hubo confusión de sonoridad alcanzan sólo un 1,58%.

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 sólo alcanzan el 4,16%.

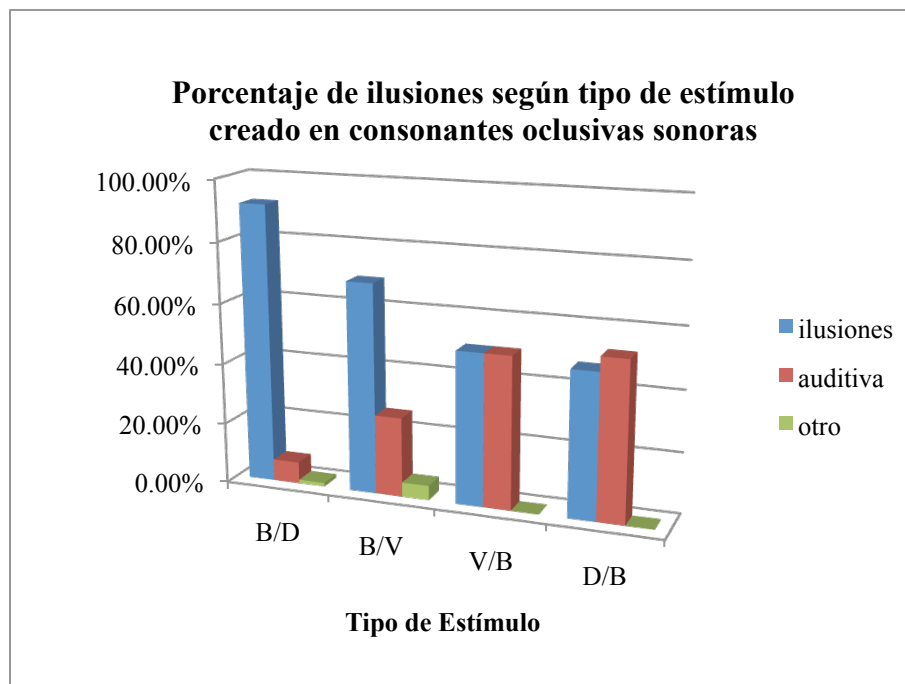


Figura 10 Gráfico de porcentajes de ilusiones por tipo de estímulo presentado en la serie de consonantes sonoras

La figura 10 grafica 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 sonoro constituido por la consonante bilabial como video y la consonante dental como audio, el porcentaje de ilusiones es de 91,66%, mientras que cuando 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%.

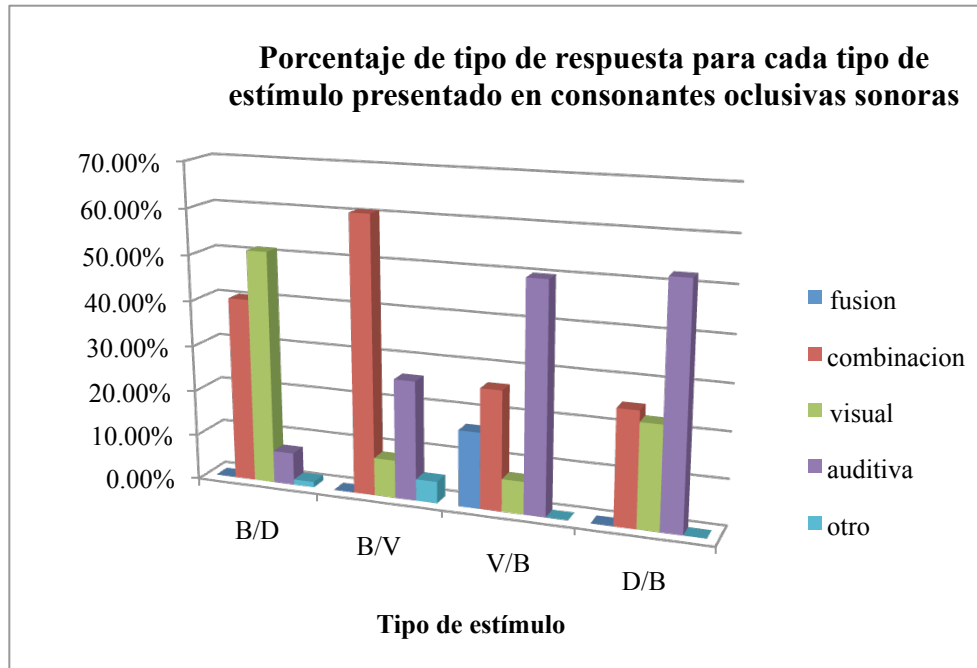


Figura 11 Gráfico de porcentajes de tipo de respuesta según tipo de estímulo presentado en la serie de consonantes sonoras

La figura 11 presenta el porcentaje de cada tipo de respuesta seleccionado por los participantes en función de los 4 estímulos presentados. Para el estímulo de 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 constituido por la misma 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 formado por una 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 constituido por la 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%.

5.3 Análisis de respuestas consonantes sordas

A continuación, se presentan los resultados obtenidos para la serie de consonantes sordas.

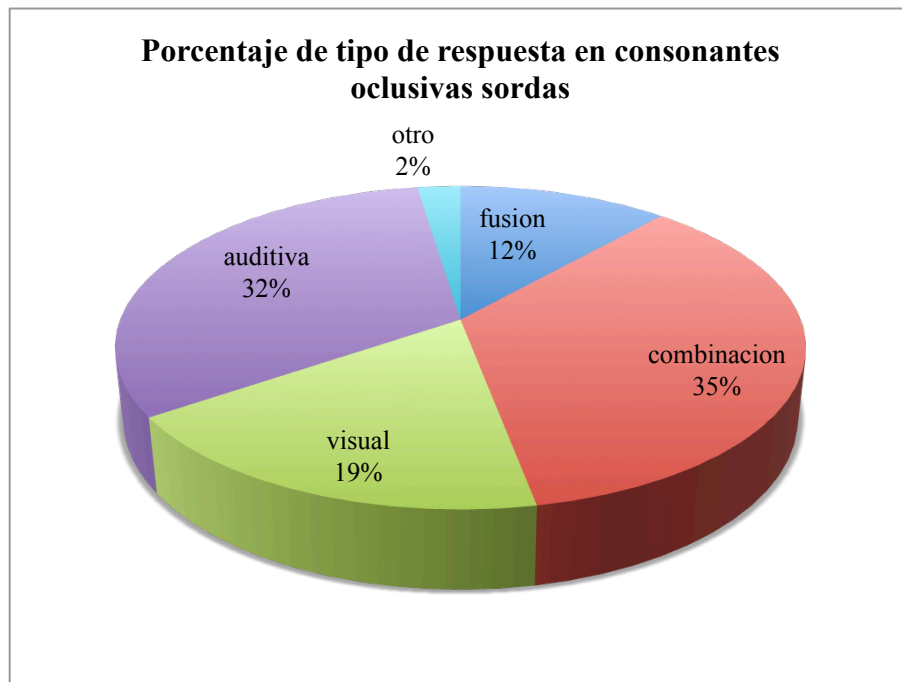


Figura 12 Gráfico de porcentajes de los tipos de respuesta seleccionados en la serie de consonantes sordas

Según los datos obtenidos, representados en la figura 12, 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%.

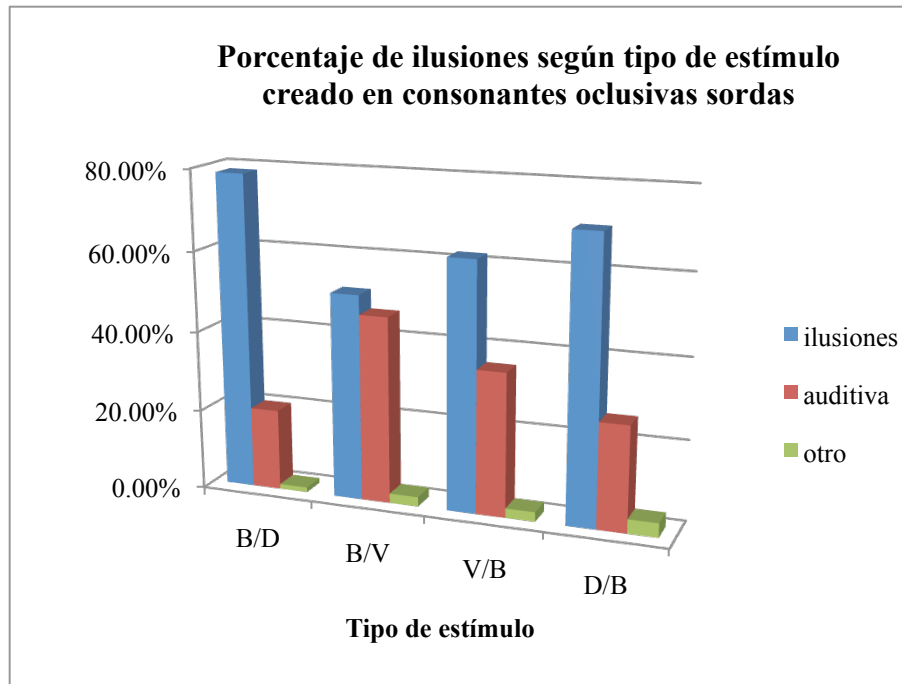


Figura 13 Gráfico de porcentajes de ilusiones por tipo de estímulo presentado en la serie de consonantes sordas

La figura 13 muestra el detalle del 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 de 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 en que la consonante dental es presentada como video y la consonante bilabial como audio, con un 70,23%. El estímulo 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 donde la consonante bilabial fue presentada como video y la velar en la 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.

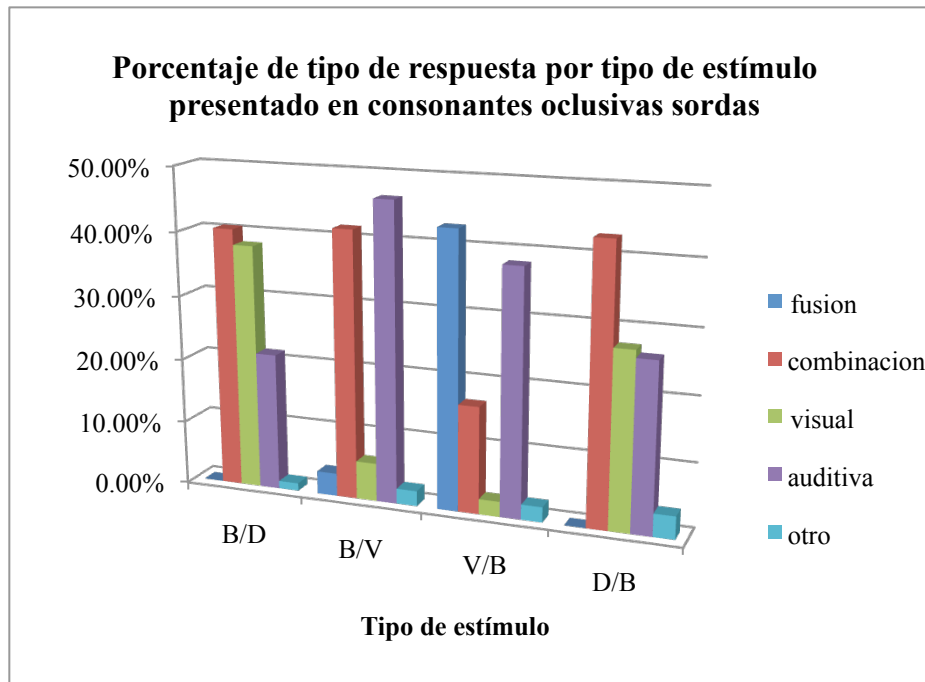


Figura 14 Gráfico de porcentajes de tipo de respuesta por tipo de estímulo presentado en la serie de consonantes oclusivas sordas

Como se observa en la figura 14, para las consonantes sordas, en el estímulo constituido por la consonante bilabial en modalidad visual y la 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%.

Cuando la consonante bilabial fue presentada como estímulo visual, pero 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 la 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 constituido por la consonante dental sorda en modalidad visual y la 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.

Luego de analizar los porcentajes obtenidos en el experimento, se presenta un análisis estadístico de las frecuencias de respuestas obtenidas mediante 2 pruebas de Chi-cuadrado.

5.4 Análisis mediante pruebas de Chi-cuadrado

Como se explicó anteriormente, se sometió 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. En la primera prueba se consideró el total de estímulos sonoros y sordos. Para esta prueba, tal como se aprecia en la figura 15, se obtuvo 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 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		

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

Figura 15 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, presentada en la figura 16, 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 H0.

Para B/V la frecuencia observada supera la esperada en las combinaciones, respuestas auditivas y otros. 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.

Tabla de contingencia tipo de estímulo-tipo de respuesta

			TipoResp					Total
			audi	comb	fusi	otro	visu	
TipoEst	B/D	Recuento	23	68	0	2	75	168
		Frecuencia esperada	55,5	61,8	13,3	3,3	34,3	168,0
		Residuo	-32,5	6,3	-13,3	-1,3	40,8	
	B/V	Recuento	61	86	3	6	12	168
		Frecuencia esperada	55,5	61,8	13,3	3,3	34,3	168,0
		Residuo	5,5	24,3	-10,3	2,8	-22,3	
	D/B	Recuento	66	57	0	3	42	168
		Frecuencia esperada	55,5	61,8	13,3	3,3	34,3	168,0
		Residuo	10,5	-4,8	-13,3	-,3	7,8	
	V/B	Recuento	72	36	50	2	8	168
		Frecuencia esperada	55,5	61,8	13,3	3,3	34,3	168,0
		Residuo	16,5	-25,8	36,8	-1,3	-26,3	
Total		Recuento	222	247	53	13	137	672
		Frecuencia esperada	222,0	247,0	53,0	13,0	137,0	672,0

Figura 16 Tabla de contingencia para tipo de estímulo y tipo de respuesta en consonantes sonoras y sordas

La segunda prueba (figura 17) se realizó usando 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		

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

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

Figura 17 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 (figura 18), 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, corresponde 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'.

Tabla de contingencia Tipo de estímulo – Tipo de respuesta con capa Tipo de consonante

TipoCons				TipoResp					Total
				audi	comb	fusi	otro	visu	
Sonora	TipoEst	B/D	Recuento	6	34	0	1	43	84
			Frecuencia esperada	28,5	32,0	3,5	1,3	18,8	84,0
			Residuo	-22,5	2,0	-3,5	-,3	24,3	
	B/V	Recuento	22	51	0	4	7	84	
		Frecuencia esperada	28,5	32,0	3,5	1,3	18,8	84,0	
		Residuo	-6,5	19,0	-3,5	2,8	-11,8		
	D/B	Recuento	44	21	0	0	19	84	
		Frecuencia esperada	28,5	32,0	3,5	1,3	18,8	84,0	
		Residuo	15,5	-11,0	-3,5	-1,3	,3		
	V/B	Recuento	42	22	14	0	6	84	
		Frecuencia esperada	28,5	32,0	3,5	1,3	18,8	84,0	
		Residuo	13,5	-10,0	10,5	-1,3	-12,8		
	Total	Recuento	114	128	14	5	75	336	
		Frecuencia esperada	114,0	128,0	14,0	5,0	75,0	336,0	
		Residuo							
Sorda	TipoEst	B/D	Recuento	17	34	0	1	32	84
			Frecuencia esperada	27,0	29,8	9,8	2,0	15,5	84,0
			Residuo	-10,0	4,3	-9,8	-1,0	16,5	
	B/V	Recuento	39	35	3	2	5	84	
		Frecuencia esperada	27,0	29,8	9,8	2,0	15,5	84,0	
		Residuo	12,0	5,3	-6,8	,0	-10,5		
	D/B	Recuento	22	36	0	3	23	84	
		Frecuencia esperada	27,0	29,8	9,8	2,0	15,5	84,0	
		Residuo	-5,0	6,3	-9,8	1,0	7,5		
	V/B	Recuento	30	14	36	2	2	84	
		Frecuencia esperada	27,0	29,8	9,8	2,0	15,5	84,0	
		Residuo	3,0	-15,8	26,3	,0	-13,5		
	Total	Recuento	108	119	39	8	62	336	
		Frecuencia esperada	108,0	119,0	39,0	8,0	62,0	336,0	

Figura 18 Tabla de contingencia para tipo de estímulo y tipo de respuesta con capas ‘sonoras’ – ‘sordas’

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.

6. Discusión

Si se busca hacer una comparación con otras investigaciones, es preciso mencionar que existen ciertas diferencias importantes en la realización de ellas. En primer lugar, tanto en Behne *et al* (2007), Colin *et al* (2002) y Gentilucci y Cattaneo (2005), los experimentos son realizados sólo con el par velar-bilabial, ya sea con el primer fonema como audio y el segundo como video o viceversa; por el contrario, en este estudio se han utilizado los pares velar-bilabial y dental-bilabial.

En segundo lugar, los dos primeros estudios mencionado, se enfocan específicamente en la variable sonoridad; es decir, en determinar si la sonoridad de la consonante tiene alguna repercusión en la percepción de los espectadores. Si bien en este estudio se han usado tanto consonantes sonoras como sordas para determinar si existen diferencias ante la presentación de unas y otras, no ha sido esta la principal motivación. Como se explicó anteriormente, esta investigación pretende dar luces en torno a la percepción de los fonemas oclusivos, considerando sus diversas características acústicas y articulatorias.

De este modo, considerando ambas series de consonantes, es posible extraer algunas ideas que den luces sobre la percepción de las mismas. Como primera aproximación y tomando en cuenta el porcentaje de respuestas, es posible hipotetizar que el alto porcentaje de ilusiones en ambos pares bilabial-dental, puede deberse a que la información acústica de la consonante dental es poco robusta, a diferencia de la información visual de la bilabial (articulación) que tendría gran relevancia para la percepción.

Al observar las respuestas de los participantes en el par bilabial-velar se puede determinar que, si bien el porcentaje de ilusiones es más alto que el de respuestas auditivas, es menor que en el caso anterior y las respuestas auditivas también aumentan, lo que podría indicar un mayor peso perceptual de la velar como componente auditivo.

En ambos casos en que la bilabial fue usada como el componente visual, las respuestas se ajustan a la teoría: la presencia de la articulación del fonema bilabial tiene un gran peso perceptual, por lo que no puede ser ignorado. En el par bilabial-dental, el mayor porcentaje de respuestas corresponde a las de tipo visual, lo que podría señalar la poca relevancia acústica de este fonema. En cambio, en el par bilabial-velar se puede observar un número más elevado de combinaciones, lo que implica

necesariamente la percepción de ambos estímulos. Esto, entonces, indicaría que la consonante velar tiene, acústicamente, un mayor peso perceptual que la dental.

Como se pudo apreciar en la figura 8, para el estímulo formado por la articulación de una consonante velar y el audio de una consonante bilabial, la mayoría de las respuestas se concentran en las que consideran sólo el aspecto auditivo (42,85%). Les siguen las de fusión (30%) y finalmente las de combinación (21,42%). Estos resultados se asemejan a los obtenidos en Behne *et al*, quienes en similares condiciones reportan un 48% de respuestas auditivas y un 34% de respuestas de fusión. Es importante mencionar, no obstante, que el número de participantes en este estudio es n=10. La lengua nativa estudiada en este caso fue el noruego.

Para el mismo tipo de estímulo, Colin *et al*, reportan en promedio un 5% de respuestas de fusión en hablantes nativos de inglés. En Gentilucci y Cattaneo, en promedio, la información acústica fue usada en un 70% de los casos; el fenómeno no fue probado sistemáticamente para los hablantes de italiano. Los autores atribuyen esto a características propias de la lengua.

Los resultados obtenidos en el estímulo dental-bilabial, en donde las respuestas auditivas alcanzan un 39,28%, las de combinación 33,92% y las visuales un 25%, indican una menor uniformidad en las respuestas de los participantes. Esto podría deberse a que la consonante dental, tanto auditiva como visualmente no tienen una presencia importante en relación a los otros dos tipos de consonantes.

Si se comparan sólo los resultados de la serie sonora con los obtenidos por Behne *et al*, se puede concluir que para el noruego el número de respuestas auditivas es casi el doble que en este estudio (61%). Las respuestas visuales son sólo de un 8%. En dicho estudio, las respuestas de fusión para la consonantes sonoras representan un 31% del total.

En el detalle de los tipos de ilusiones generados por la serie sonora, presentados en la figura 11, se pudo observar que en el estímulo bilabial-dental, 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%). Como se mencionó anteriormente, 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 velar-bilabial, el porcentaje de respuestas auditivas fue de 50%. mientras que el de las de fusión fue 16,66% y de las de combinación 26,19%. En los estudios citados, 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 superó el 17% en el caso de las consonantes sonoras para el mismo estímulo.

En los estudios consultados 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 alcanza 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, 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 la figura 14.

En el detalle de las consonante sordas se puede apreciar que en el estímulo velar-bilabial, 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 (consonante bilabial-consonante velar), las combinaciones en la serie sonora superan en un 20% a la serie sorda.

7. Conclusiones

Al realizar el experimento dentro del paradigma McGurk, se ha podido 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 respondieron 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 ha podido 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 cumplieron los tipos de ilusiones esperados según los tipos de estímulos presentados. Al observar la tabla de contingencia 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 superó la esperada. Tanto para B/V, como para V/B, la interacción con los tipos de respuesta de combinación y fusión, respectivamente, mostró ser estadísticamente significativa. Más aún, la presente investigación permitió 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 que 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 pudo 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 (según lo que informan investigaciones anteriores). 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 que tendrían las consonantes oclusivas sordas, es precisamente 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 esa 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 resolver o intentar dar una solución más certera a los cuestionamiento 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. Además, como se mencionó al comienzo de esta investigación, a través de un paradigma McGurk es posible aportar en diversas áreas de las ciencias humanas. Es por ello que resulta fundamental continuar con este tipo de investigaciones y crear nuevos diseños experimentales que permitan desarrollar todo el potencial que guarda el efecto McGurk.

8. 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". Obtenido el 27 de enero de 2012, de <http://www.fon.hum.uva.nl/paul/papers/McGurk.pdf>
- Brancazio, L., Best, C. y Fowler, C. (2006). "Visual influences on perception of speech and nonspeech voal-tract events". *Language and speech* 49, 21-53.
- Campbell, R. (1992). "The neuropsychology of lipreading". *Biological Sciences* 335, 39-45.
- 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.
- Gósy, M. (1992). *Speech Perception*. Frankfurt: Hector.
- Hardison, D. (2007). "The visual element in phonological perception and learning". En M. C. Pennington (Ed.), *Phonology in context* (pp. 135-158). New York: Palgrave Macmillan.
- Lieberman, P. Y Blumstein, S. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.
- Martínez, E. y Fernández, A. (2007). *Manual de fonética española*. Barcelona: Ariel
- Massaro, D. Y Cohen, M. (1990). "Perception of synthesized audible and visible speech" *Psychological Science*, 1.
- Mathe, C. (2002). "Fuzzy Logic". Obtenido el 27 de enero de 2012 , de www.aic.ac/en/books/Fuzzy%20Logic.doc

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 2003.

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.

Soto Barba, J. (1994). "¿Los fonemas /b/ y /p/ se diferencian por la sonoridad?" *Estudios filológicos* 29, 33 – 38

Soto Barba, J. y Valdivieso, H. (1999). "Caracterización fonético-acústica de la serie de consonantes /p,t,k/ vs. /b,d,g/" *Onomázein* 4, 125-133.

Tatham, M. Y Morton, K. (2006). *Speech production and perception*. Hampshire: Palgrave Macmillan