



窺探視聽盛宴的背後

——「視聽整合：從知覺到感受」[#]

葉素玲*

一、前言

我們生活的世界裡充滿了五彩色光與萬籟聲響，人們無時無刻不在享受這一場視與聽的盛宴。正如美味的食物需要五味調和一樣，視聽訊息只有協調的整合在一起，才能形成和諧的感覺體驗，獲得對外界的正確認知。

視覺訊息和聽覺訊息的整合對於人類而言似乎是一件理所當然的事，我們只要一睜開眼睛，就能感受到這個完整的世界，毫不費力。然而，這個整合過程其實需要解決一系列的計算難題。

諸多難題中最主要的是時間整合。人類通過不同的感官通道來處理色光與聲音：光線透過眼球，入射到視網膜上，轉換成神經信號，上傳至視覺皮層處理得到視知覺；聲音則引發鼓膜的震動，藉由聽小骨傳遞給耳蝸形成神經信號，再傳入聽覺皮層形成聽覺。兩條通路處理資訊的速度並不相同，一般情況下，聽覺訊息的處理速度較快。而且由於雜訊、物理刺激本身強度等因素，各種具體情況下資訊處理的速度均不相同。因此，同一時刻發生的視覺事件和聽覺事件，大腦在不同時刻處理完成，且二者的時間差不恆定，為整合視覺事件和聽覺事件帶來了困難。工程師在解決類似問題時，往往會給兩個速度不同的系統打上附加的時間標籤，用以同步二者處理的資訊。然而視覺和聽覺系統所接收的資訊顯然不包含統一的時間標籤。因此，人類認知系統如何進行時間整合就成為了心理學研究的熱門問題之一。經過多年探索，目前普遍觀點認為，知覺系統存在整合時間窗以容忍時間誤差，該時間窗內的視聽訊息會整合成為單一知覺事件。隨著研究的深入，對整合時間窗的認識也逐步加深，例如時間窗的大小會隨年齡變化、對於不同層級的知覺

[#] 本文內容為作者執行國科會補助計畫「視聽整合：從知覺到感受」(2009 ~ 2012)研究成果。

* 國立臺灣大學心理學系教授兼系主任

單元，會有不一樣大小的時間窗等等。

另一個主要難題是空間整合問題。整合不僅包括將統一時間點的事件整合入一個知覺單元，也包括將同一地點發生的事件整合到一起。人類的視敏度較高，相對的，聽覺分辨方位的能力就弱很多，我們只能大致判斷聲音來源的方向。因此，複雜環境中如何將聲音與發聲物進行正確的整合就成為一個難題。與時間整合相同，空間整合也存在誤差容許的範圍。在這個範圍內發生的視聽覺事件將會整合到一起，甚至會反過來影響知覺到的視覺和聽覺刺激的位置。

研究還發現，認知因素在知覺整合中也很重要。知覺整合並非機械的將一定時空範圍內的事件進行整合，而是根據所獲得的資訊，主動的去構建知覺表徵。該構建過程考慮到視覺事件和聽覺事件本身的意義及其相互關係，用以作為額外線索，彌補知覺整合資訊不足的問題（視覺和聽覺資訊本身不足以確定二者的對應關係）。Welch 與 Warren (1980) 指出，認知因素包含以下三項：是否察覺感官間刺激的結構、對不同感官刺激的整體性假設 (unity assumption)，以及感官間刺激共同出現的必然性 (compellingness)。換句話說，認知系統會根據所掌握的因果結構知識，在諸多可能中選擇概率較高、具有整體性的整合方式來形成知覺輸出。以腹語術為例，當演員操縱木偶活動嘴部，同時在自己嘴不動的情況下為其配音，觀眾聽到了說話聲，也看到嘴不動的演員與嘴活動的木偶。此時有諸多可能的整合方式，例如：(1) 木偶在說話；(2) 木偶只是動嘴巴，而聲音是由嘴不動的演員發出的；(3) 木偶只是動嘴巴，聲音是從別的地方傳來的等等。由日常生活經驗可知，說話往往伴隨著嘴部運動，二者具有整體性，光張嘴不出聲和光出聲不張嘴的概率相對較小。比較上述三者的聯合概率，便可發現方式 (1) 的概率最高，方式 (2) 結合了兩種概率較低事件，而方式 (3) 則需要假設外部傳來的聲音恰好與木偶的動作同步，概率更低。因此我們的認知系統選擇了概率最高、最具一致性的方式 (1) 來整合視聽訊息形成知覺，從而感受到木偶發聲的奇妙感覺。所以，知覺整合實質上是認知系統利用從外部世界獲得的極其有限的資訊，根據事件的意義和先驗知識構建一個結構化的內部知覺世界的過程。

過去在探討知覺議題時，限於所知及研究方法的限制，使得研究者必須先將研究範圍單純化，因此多著重在單一感官的訊息處理。近年來相關理論不斷發展，實驗技術日趨成熟，為複雜刺激的視聽整合研究提供了堅實基

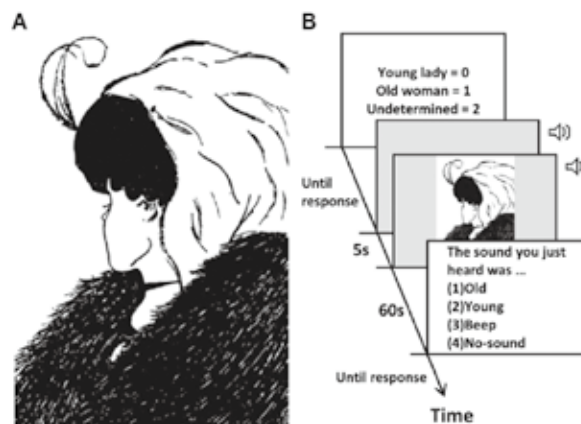


礎。因此，本實驗室延伸過去針對單純刺激（如光點、純音）所探討的視聽整合知覺，採用具有生態效度的複雜刺激（如自然場景、音樂）進行國科會計畫「視聽整合：從知覺到感受」支持下的一系列研究，藉此探討認知因素（如影響整體性假設的一致性原則）對視聽整合的影響。本文將對其中部分研究進行簡要介紹，以期促進相關學術交流。

二、雙穩定圖形與視聽整合

雙穩定圖形是指可以有兩種不同知覺結果的圖形，例如圖一（A），圖形既可以看成一個年輕女孩，也可以看作是一個老年婦人。雙穩定圖形具有能長時間呈現且一次只能知覺到一個物件的特性，非常適合作為探討視覺意識的刺激材料。如果視聽整合過程中，認知系統會利用語音的年齡特徵，根據整體性原則將視聽訊息整合成協調的知覺整體的話，那麼我們可以預期，當播放年輕女孩聲音時，受試者將有更多知覺到女孩的圖形；而播放老婦聲音時，將更傾向於知覺到老婦的圖形。因此我們設計了實驗檢驗上述假設。實驗過程如圖一（B）所示，雙穩定圖形配合聲音呈現，受試者觀察圖形一分鐘並按鍵報告所看到的是女孩還是老婦，結束後回答聽到何種聲音。

實驗結果顯示對於雙穩定圖形的意識知覺會受到語意訊息的促發，即聽到年輕女人的聲音時，會促進雙穩定圖形中對於年輕女人的意識知覺，反之亦然（Hsiao et al., 2012）。因此，複雜刺激的視覺整合所利用的線索不僅包括



圖一（A）老女人與年輕女人的雙穩定圖像。（B）實驗會先提示受試者要做的作業，接著先出現五秒鐘的聲音，隨後是一分鐘的圖像呈現。受試者於一分鐘當中，按鍵表示知覺狀態，一分鐘結束後，受試者回答聽到什麼聲音

視聽刺激的物理特徵，甚至還上升到諸如人物年齡等深層語意層次，足可見人類認知系統的智慧化程度。而且，這些複雜運算只在接受到刺激後的一瞬間便可完成，自動形成意識內容，其高效性令所有人工智慧圖像處理系統都望塵莫及。由此可見，深入研究認知系統該方面的演算法特性，透澈理解視覺整合的內部機制，將大大促進相關計算模型的優化，提升人工系統的智慧化程度。

三、臉部訊息的隱式處理與視聽整合

前述實驗證明人們覺察到的視聽資訊的整合可以受到語意的觸發，那麼這種促進作用是否可以延伸到意識閾限之下的刺激呢？為了探索該問題，我們進行了臉部訊息的隱式處理與視聽整合的系列實驗，探討可否藉由相配的聲音而促進臉部訊息的意識觸接。該系列實驗採用了持續閃現抑制派典（continuous flash suppression, CFS）產生闕下臉部訊息。實驗中，我們於受試者的一隻眼睛呈現不斷閃現的高對比度圖塊，而於另一眼呈現低對比度的臉部圖片。由於雙眼競爭，受試者只能看見閃動的圖塊，而臉部圖片被壓抑在意識閾限之下。觀看一段時間後，壓抑的刺激將有機會進入視覺意識，視覺刺激從雙眼抑制中偵測所需的反應時間即可作為隱式處理的指標。我們採用3D 虛擬軟體 FaceGen Modeller 3.31（Singular Inversions, Toronto, Canada）排除可能產生的低階訊息影響，產生出人臉圖片呈現於 CFS 條件下，同時配上人聲或嗶聲。受試者被要求在看到人臉時盡快按鍵報告，並記錄反應時間，該反應時間反映了人臉圖片從抑制中解放所需的時間。

結果顯示，臉孔正立呈現情況下，唯有人聲相較於無聲情境可促進臉部偵測；倒立情況下，人聲與嗶聲的效果相同。我們推論正立的人臉因為熟悉度的關係，必須要有意義的聲音才有促進的效果；然而倒立的臉因為不熟悉，所以只要有聲音都有提高警覺的作用。

在此基礎上，為了進一步探討意義一致性對臉部視聽整合的影響，我們操弄臉部表情的正負向及聲音語調情緒的正負向情緒，探討情緒一致性是否可以促進臉部整合。結果是肯定的。此結果顯示臉部訊息的視聽整合不僅局限於人臉與人聲的一致性，也可達到情緒類別的一致性。

該系列實驗由於採用闕下刺激，可避免高層認知過程的影響，進一步確認了認知系統在視聽整合過程中可以自動利用語意一致性線索。闕上刺激由



於受試者可以覺察到刺激本身並聯想到視聽刺激間的關係，有可能猜測到實驗的真實目的或採用一些特殊策略，從而改變自己回答的傾向。而當受試者根本意識不到刺激存在時，便不可能進行與此相關的高層思維了。因此實驗結果能更好的反映認知系統真實的特性。

四、自然場景的認知因素與視聽整合

上述研究表明語意一致性可以促進聲音與視覺物體的整合；然而人們往往處於視覺訊息的洪流中，所需要處理的不僅是單獨的物體，而且需要面對包含了許多物體的複雜場景。那麼在場景知覺中，語意一致性是否也有同樣的效應呢？我們操弄視聽雙感官訊息的場景刺激材料（餐廳與街道），使用前述之連續閃爍抑制派典探討在隱式歷程下視聽訊息的整體一致性是否會加速對場景的偵測。結果顯示，當聲音與視覺場景一致時，即街道場景配上車輛行駛聲，餐廳場景配上餐具敲擊聲，壓抑下的視覺場景將更快被偵測到。該結果表明語意一致性對視聽整合的效應可以推廣至複雜的視覺場景。

五、結論與未來展望

綜上所述，我們系統性的探究了雙穩定圖形、臉孔辨識、場景知覺過程中的視聽整合，結果均發現語意一致性可促進視聽整合。這些結果進一步支持了認知因素在知覺整合過程中的重要作用，表明知覺整合是積極主動根據刺激意義建構統合知覺表徵的過程。也就是說，我們的認知系統才是這一場場視聽盛宴的真正締造者。

Bedford (2001) 藉由空間、時間、跨感官及雙眼訊息等不同因素探討物體辨識，並認為這是一個對待整合刺激進行同一性決策 (identity decision) 的歷程。Vatakis 與 Spence (2007) 延伸這樣的想法，將跨感官的整合分為兩類，一類為需要刺激導引 (stimuli-driven) 的因素，例如是否在相近的時間、空間出現。另一類則為認知 (conceptual-driven) 因素，即在概念上，我們是否會假設其為單一而整體的事件，這樣的整體性假設有時可助於提高刺激層次在時間空間不同步的相容性。我們的系列研究進一步支持並拓展了上述理論。

基於本實驗室多年來對於視聽整合方面的研究經驗，本計畫還進一步拓展到複雜且跨感官的知覺研究，內容創新且具挑戰性。例如，我們的研究拓展到音樂與影像的情緒評估、視聽整合的時間感知覺、喜好偏好的認知與情

緒因素、包括嗅、味、觸、視、聽的跨感官整合等議題在跨感官整合上的表現。我們發現：當影音兩者的情緒意涵一致時可增強所知覺到的情緒強度；時間感知覺同時受到視聽訊息的影響，其中視聽訊息的變異度影響兩者感官訊息的權重；對於熟悉事物的偏好倚賴知覺歷程處理的流暢度，對於新奇事物的偏好倚賴注意力處理歷程；主觀熟悉感與喜好感受受到不同情緒的調節；不同感官（嗅、味、觸、視、聽）之間的整合與交互作用。此一系列的實驗，不僅將知覺整合研究拓展至新的領域及更高的認知層次，而且對其他感官如觸覺、嗅覺研究、注意力研究和視聽整合辨識提升現象的背後機制，具有一定的啟發性。

藉由我們的視聽整合研究及前人的相關工作，我們對多重感官整合已經有了初步了解；然而，要更深入全面的了解人類複雜交錯的知覺系統、甚至於情緒與偏好等感受方面等議題，還需要更多的研究來探討。我們認為，進一步的研究可以從以下幾個方面來展開：

1. 拓展認知因素研究的廣度，進一步探索可被大腦利用的認知線索，檢視這些線索在不同感官系統中的一致性對整合過程的影響及其交互作用。
2. 研究知覺整合的發展過程，了解整合機制的發生與發展規律，探討認知因素中有關知識結構是如何形成的。
3. 完善知覺整合理論，並引入相關數學工具，構建整合過程的認知計算模型，用以模擬人類知覺整合行為，為工程應用領域類似問題的解決提供參考。

透過上述三個方面的研究，可以繼續深化知覺整合研究，並推動不同學科領域的交流與合作，並藉此促進心理學理論的發展與心理學應用實踐的推廣。

參考文獻

- Bedford, F. (2001). Towards a general law of numerical/object identity. *Current Psychology of Cognition*, 20 (3/4), 113-176.
- Hsiao J. Y., Chen Y. C., Spence C., & Yeh, S. L. (2012). Assessing the effects of audiovisual semantic congruency on the perception of a bistable figure. *Consciousness and Cognition*, 21 (2), 775-787.
- Vatakis, A., & Spence, C. (2007). Crossmodal binding: evaluating the “unity assumption” using audiovisual speech stimuli. *Perception and Psychophysics*, 69 (5), 744-756.
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88 (3), 638-667.